UNIVERSITÉ DE MONTRÉAL

CONTENT-BASED IMAGE ENHANCEMENT FOR EYE FUNDUS IMAGES VISUALIZATION

MAXIME SCHMITT
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
JUILLET 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

CONTENT-BASED IMAGE ENHANCEMENT FOR EYE FUNDUS IMAGES VISUALIZATION

présenté par: SCHMITT Maxime
en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées
a été dûment accepté par le jury d'examen constitué de:

M. LANGLOIS J.M. Pierre, Ph. D., président
M. HURTUT Thomas, Ph. D., membre et directeur de recherche
Mme CHERIET Farida, Ph. D., membre

DEDICATION

To my parents,
For their never ending support

RÉSUMÉ

Un nombre croissant de maladies peuvent être détectées par l'analyse de la rétine humaine, et ce même à des stades de celles-ci où le patient pourrait ne pas encore avoir de symptômes visibles. Comme les traitements sont généralement plus intrusifs et coûteux pour les stades les plus avancés des maladies, il est préférable pour sa santé et ses finances que le patient soit soigné au plus tôt. C'est pourquoi le dépistage régulier de la population est considéré comme un des moyens les plus efficaces pour limiter le nombre de cas extrêmes. Étant donnée la quantité d'examens de la rétine que l'on voudrait alors effectuer chaque année, l'amélioration du processus de dépistage est un sujet important. Les propositions à ce sujet cherchent généralement soit à automatiser l'analyse, soit à aider les médecins à faire leur diagnostic.

Ce mémoire propose d'améliorer le processus de dépistage grâce à une méthode d'amélioration des images de fond d'œil pour faciliter leur visualisation et leur diagnostic par un examinateur humain. Les principaux problèmes que peuvent présenter les images de fond d'œil et dont la correction peut aider à simplifier leur diagnostic sont liés à du contenu flou, des artefacts de réflexions, des défauts d'illumination et de contraste ou encore à une importante variabilité de taille, de forme ou de couleurs entre les images. L'objectif de notre travail est de proposer une méthode qui permette de mieux observer les éléments présents dans ces images afin de faciliter leur analyse tout en s'assurant que l'apparence des images reste plausible, en particulier en termes de couleurs, de telle sorte que l'examinateur ne soit pas gêné par les changements. Nous cherchons également à réduire la variabilité en couleurs entre les images en leur faisant toutes partager la même palette de couleurs.

Les précédents travaux portant sur l'amélioration des images de fond d'œil se concentrent sur la correction des artefacts dans les images comme le flou ou les problèmes d'illumination et de contraste non uniformes. Bien que les méthodes proposées améliorent la visibilité des éléments dans les images, elles ne sont généralement pas adaptées pour la visualisation par un examinateur mais plutôt pour être utilisées comme pré-traitement pour d'autres méthodes automatiques qui peuvent mettre à profit ces améliorations. En effet, l'apparence de leurs résultats sur les images couleurs a tendance à être trop différente des images habituelles et les examinateurs ont donc des difficultés à travailler sur les images produites. L'un des principaux problèmes est la disparition de certains éléments des images lors du traitement, comme par exemple la fovéa, qui peuvent ne pas être nécessaires pour certains algorithmes automatiques mais qui sont des points de repère importants pour les examinateurs. De plus, les couleurs des images produites ne sont pas naturelles, ce qui gêne les examinateurs pour faire leur

diagnostic. Nous nous intéressons alors à un autre domaine d'amélioration d'images au travers des techniques de transfert de couleurs par l'exemple qui modifient une image source pour qu'elle utilise les couleurs d'une image de référence. Cette approche est généralement utilisée pour le transfert de style ou d'ambiance entre des images naturelles et n'a pas encore été appliquée aux images de fond d'œil. Des travaux récents proposent d'utiliser le contenu des images pour guider le transfert de couleurs et présentent donc des résultats dans lesquels les différents éléments partagent les couleurs des éléments correspondants dans l'image de référence. Comme notre objectif est de modifier uniquement l'apparence des images et non leur contenu, l'usage d'une méthode de modification des couleurs est approprié.

Ce travail propose d'étendre et d'adapter au contexte des images de fond d'œil une méthode de transfert de couleurs qui utilise les textures dans les images pour guider le transfert. L'algorithme original utilise une valeur de similarité entre les pixels calculée à l'aide d'un descripteur de textures pour guider le transfert de couleurs. Les images ainsi produites ont des couleurs proches de l'image de référence mais sont globalement trop sombres et n'utilisent pas les couleurs attendues pour le réseau vasculaire. Pour résoudre le premier problème, nous proposons d'utiliser une segmentation de la région d'intérêt pour ne plus utiliser les pixels noirs en dehors du champ de l'appareil de capture lors du transfert. La teinte globale des résultats est alors plus proche de l'image de référence mais les couleurs du réseau vasculaire restent trop différentes de la référence. C'est pourquoi nous proposons de modifier la mesure de similarité entre les pixels pour utiliser une segmentation des réseaux vasculaires afin d'augmenter la similarité entre les pixels des réseaux. Les vaisseaux dans nos résultats ont alors des couleurs plus proches de ceux de l'image de référence.

Nous présentons une expérience avec des ophtalmologistes dans le but de mesurer l'effet de notre méthode sur leurs performances en termes de vitesse et de précision de diagnostic. Bien que les résultats de notre expérience montrent une accélération pour un des médecins avec les images améliorées par notre méthode, nous ne pouvons pas conclure que cette différence est le fruit de nos améliorations à cause de certains biais liés à notre protocole. Cela nous a cependant permis d'identifier les limites et les biais de notre protocole qui devront être pris en considération pour une éventuelle prochaine itération de celui-ci.

La méthode proposée remplit les objectifs fixés d'augmenter la lisibilité des images et de réduire la variabilité en couleurs entre elles. De futurs travaux s'intéressant à l'utilisation d'un descripteur à plusieurs échelles pourraient permettre d'améliorer encore le transfert de couleurs, en particulier pour les plus petits éléments dans les images pour lesquels le descripteur actuel n'est pas toujours adapté. Pour améliorer le processus de dépistage, de nouvelles approches de visualisation peuvent aussi être considérées comme la mise en valeur

de régions de l'image pour guider l'œil de l'examinateur vers les zones où il est le plus probable de trouver des éléments utiles pour son diagnostic.

ABSTRACT

Analyzing a patient's retina allows to check for an increasing number of diseases and conditions, especially at their early stages when the patient may not notice any symptoms. Usually, the sooner a disease is treated the better it is, both for the patient's health as well as their finances, as treatments tend to become more intrusive and costly the more advanced the condition is. For these reasons, regular screening of the population is a common recommendation to reduce the number of extreme cases. Considering the objective to test the maximum number of people each year, improving the screening process is important. The improvements proposed usually either automate the diagnosis or help the graders in making their diagnosis.

This thesis focuses on improving the screening process by proposing an image enhancement method for eye fundus photography images visualization to make it easier for the graders to make their diagnosis. Eye fundus images present many problems whose correction could help in making their diagnosis easier such as blurred content, reflection artifacts, non-uniform luminosity and contrast as well as variability in size, shape and colors. With our method, we want to make the elements in the images more visible to facilitate their localization and recognition while maintaining a plausible appearance for the whole image, especially in terms of colors, so as to not confuse the grader. We also want to have all the images using the same color palette to reduce the variability among images.

Previous works for the enhancement of eye fundus images have focused on correcting artifacts such as blur or luminosity and contrast issues. While these methods do bring improvement to the visibility of the elements in the images, they are usually better suited as preprocessing steps to other automated methods that leverage these enhancements to improve their own results. Indeed, they fail at maintaining the natural appearance of the color images and produce results that are difficult to analyze for a human expert as they do not share enough visual resemblance with regular images. In particular, they tend to make some parts of the image disappear such as the fovea which may not be necessary for some automated algorithms but that is used as a landmark by the graders. Also the colors of the resulting images are unnatural which bother the graders when they make their diagnosis. We then consider another field of image enhancement in the by-example color manipulation methods that allow to change a source image to use the colors of a given target image. This approach is usually used to change the tone or the style of regular photographs and has not yet been adapted to the context of eye fundus images. Recent works propose methods that transfer

the colors differently depending on the content and thus produce results where elements in the resulting image share the colors of the corresponding elements in the target image. As we want to modify only the appearance and not the content of the images, using a color manipulation method is appropriate.

This work expands on a color transfer method that uses the textural content in the images to guide the color transfer and adapts it to the context of eye fundus images. The original algorithm computes a similarity metric for each pixel based on a texture descriptor and uses this metric to guide the color transfer. This produces results that have colors close to the target but that are generally too dark and for which the vascular network in particular does not have the expected colors. In order to solve the first issue, we propose to use a Region Of Interest (ROI) segmentation to not take into account the black pixels outside of the camera field of view when applying the transfer. This improves the global tone of the results that is then closer to the target but the colors of the vessels still are not close enough to those of the target. We then propose a modification of the similarity metric with a segmentation of the vascular network to make pixels from the vessels to be considered more similar. This modification allows for the vessels in our results to have colors closer to those of the vessels of the target.

We conduct a user study with ophthalmologists in order to measure the effect of our method on the performances of grading in terms of speed and precision. While the experiment shows an increased grading speed for one of the ophthalmologists with our enhanced images, we cannot conclude on the impact of our method on the grading performances as some biases related to our protocol prevent us from being sure of the origin of this acceleration. This however allows us to identify the limits and biases of our protocol that should be taken into consideration for a potential future iteration of the user study.

The proposed method reaches both the objectives of enhancing the readability of the images and of reducing the variability in colors among images. Using a multiple scales descriptor could improve the color transfer on the smallest elements in the images for which the current descriptor is not always adapted. In order to further improve the screening process, other visualization methods could be considered such as region highlighting to guide the grader's eyes to suspect areas in the images.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AHE	Adaptive Histogram Equalization
CAD	Computer-Aided Diagnosis
CLAHE	Contrast Limited Adaptive Histogram Equalization
DWT	Discrete Wavelet Transform
MD	Mahalanobis Distance
MSLD	Multi-Scale Line Detection
ROI	Region Of Interest

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Diabetic retinopathy is a disease related to diabetes. It is the main cause of visual impairment and blindness among working-age adults [Nentwich and Ulbig (2015)]. This assessment is not set to change in the near future as recent studies estimate an increased number of diabetes patients of at least 20% until 2030 [Shaw et al. (2010)]. Some treatments do reduce the risk of blindness but they are both expensive and dependent on the timing at which it is provided. These considerations explain the focus on the screening process that has proven successful in reducing the number of extreme cases of the disease where it has been established [Stefánsson et al. (2000)]. Retinal photography is often used in this process. Scientific research to either assist or automate the screening thanks to image processing is therefore very active. The current process for manual diagnosis is fastidious as experts have to discern details related to the disease in images that have a lot of variations. This thesis aims at proposing a new method to enhance eye fundus images to improve this manual diagnosis process.

In this chapter, we will present the fundamentals of the eye fundus and its photography as well as the screening process. We will then introduce the roles of image processing on retinal images before presenting the layout of this thesis.

1.1 Retinal pathologies screening

1.1.1 Eye fundus photography

Eye fundus imaging gathers a large array of methods to represent the retina into two or three dimensional images such as fundus fluorescence, ultrasonography or optical coherence tomography [Abràmoff et al. (2010)]. Among the existing imaging technologies, this work is especially interested in eye fundus photography which produces 2D color images. This technology is largely used for the screening process as it is non-invasive for the patient and is fast to execute. The anatomical elements of the eye such as the optic disc or the retinal vascular vessels can be seen on the resulting image (Figure 1.1).

There is some variability among images obtained using this method, depending on the model of camera used for the capture, the intensity of the flash or some properties of the eye of the patient: each person's retina has a given color, the age of the patient as well as some conditions such as cataract can also influence the final result. Some results that show this diversity can be seen in Figure 1.2. Such variability makes the work of an expert more difficult for multiple reasons. Details that can be related to a disease are harder to discern on low

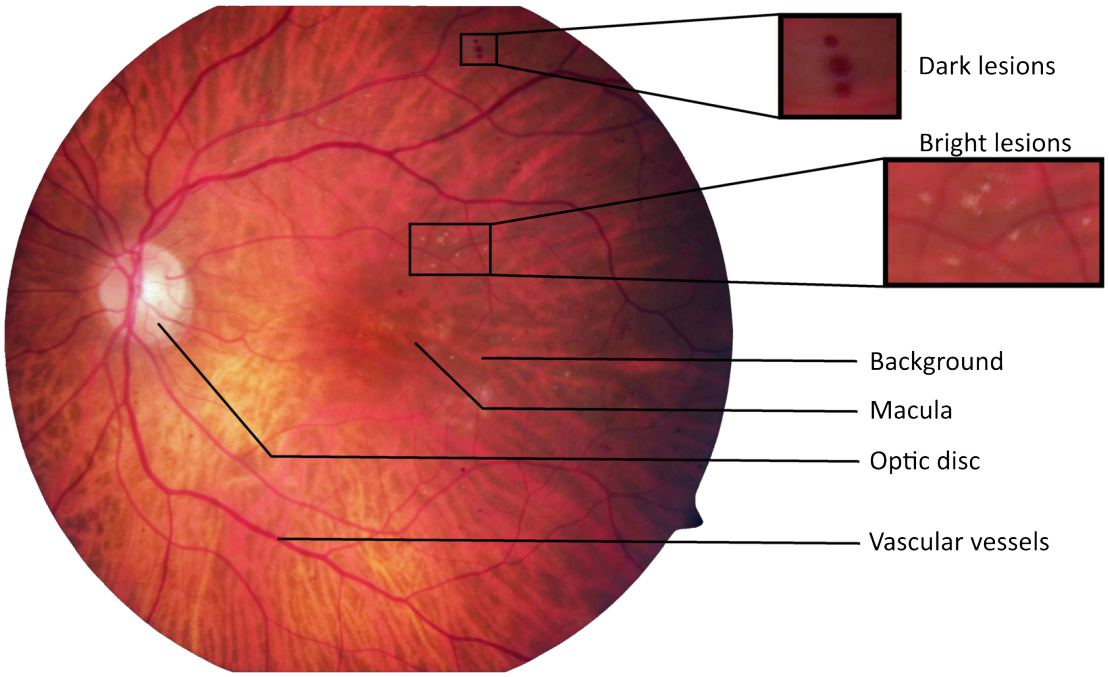


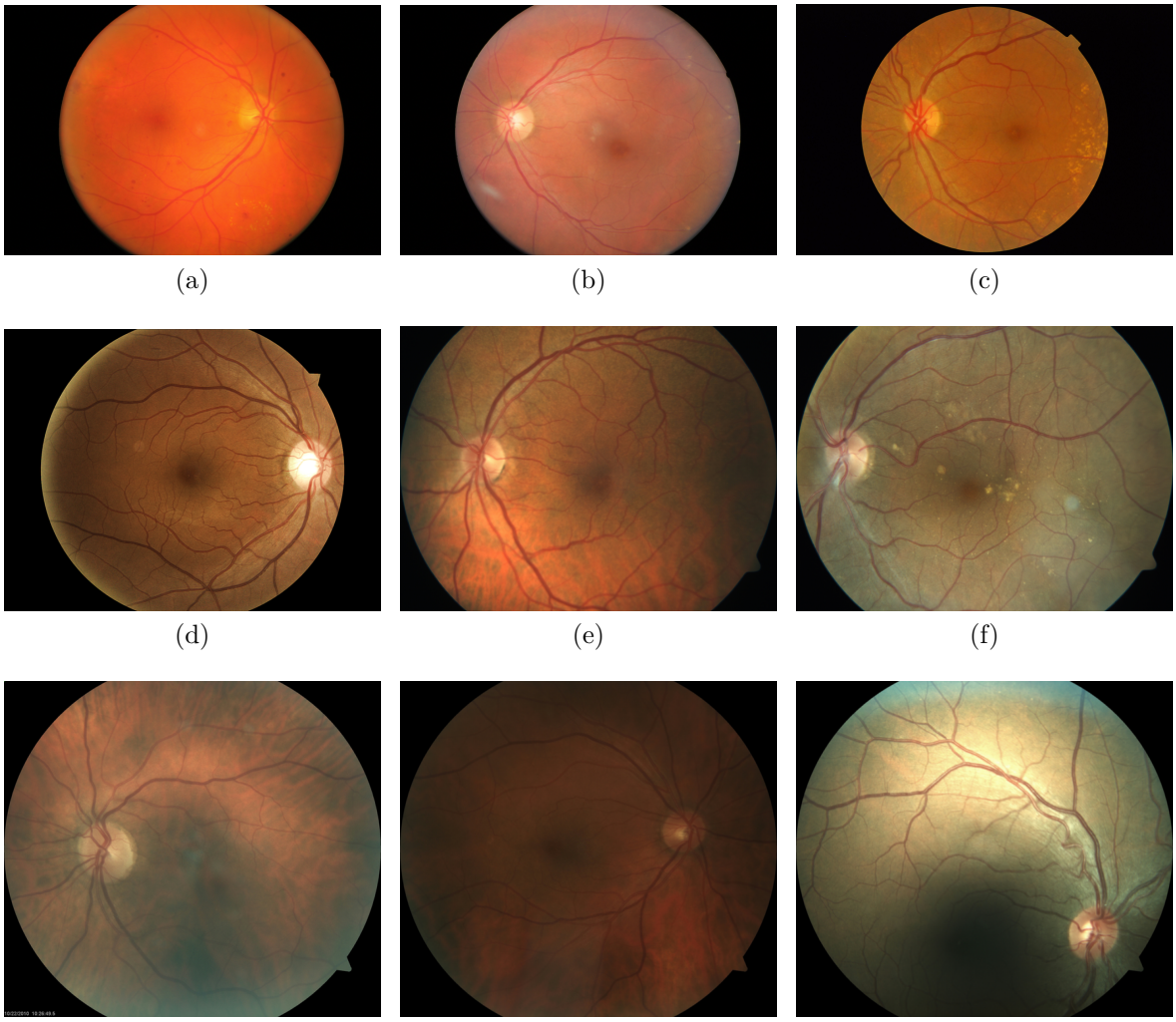
Figure 1.1 Annotated eye fundus photography (Image courtesy of Diagnos Inc.)

quality images. Also variations in size, color and shape force the expert to accommodate to each image individually, drawing no advantage from their seeming resemblance.

1.1.2 Screening process

Protocols for the screening of eye diseases can be divided in two categories: direct screening and photographic screening [Stefánsson et al. (2000)]. For a direct screening, an ophthalmologist is directly examining the eye of the patient while for the photographic screening a photographer takes an eye fundus photography of the patient that will later be examined by an expert. Those experts, that we will call graders, usually see a few dozens images every day for which they have to give a first diagnosis and recommend the patient to see an ophthalmologist in the relevant cases. While in the first implementations of photographic screening, the photographs used to be printed on films, this practice has then been replaced by digital images viewed on a computer.

While the direct screening is still widely used, the use of photographic screening is growing to optimize the screening so that more people can be examined, especially in remote areas where there can be no ophthalmologist. On the other hand, as the timing is an important factor for the treatment, it is required that this process remains efficient. Therefore, there have been many research works on improving its performance by either helping the grader in



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 1.2 Selection of images showing the important variability in size, quality, color and illumination among eye fundus images. (Images courtesy of Dr. Chakor)

making the diagnosis or even automatically giving a diagnosis. The methods resulting from those researches are generally described as Computer-Aided Diagnosis (CAD) systems.

1.2 Image enhancement for eye fundus images

In this thesis, we propose a new method to enhance eye fundus color images for the photographic screening process by modifying the colors of the original image to reduce the part of the variability between the images that is not interesting for the experts, thus helping them make their diagnosis. It is part of the vast field of research of image enhancement. We can separate image enhancement techniques into two categories depending on their purpose. The

first category is comprised of methods that aim at helping human viewers to interpret information in an image. We will refer to methods from this category as visualization methods. The second category aims at providing an image that would be a better input for a subsequent algorithm such as a vascular network segmentation method. We will refer to methods from this category as preprocessing methods.

In the case of medical imaging, there is an additional constraint: the elements in the images should not be altered in a way that would change the diagnosis whether it is done by a computer or a human expert. In this case, the main purpose of image enhancement methods is usually to remove artifacts related to the capturing device such as noise or luminosity problems. The objective is then to get an image closer to the ideal image than we would get from an ideal capturing device. This is a difficult problem because it first requires to understand the nature of these artifacts to be able to distinguish them from normal elements of the image and then to devise a method that can correct them without damaging any of those elements, since that could then change the diagnostic of the patient.

Recent research on image enhancement for eye fundus photography focuses on preprocessing methods. They are used to improve the results of other advanced methods such as retinal vessels segmentation or CAD algorithms. They often limit themselves to the green channel of the images which is widely recognized as the one where the important elements are the most visible among the three existing channels. Visualization methods gather few previous works. In particular, modifying the color image to help the graders make their diagnosis is an alternative approach that has, to our knowledge, not yet been considered. While algorithms for automatic diagnosis could potentially bring better screening performance, the current process relies on the graders and on their capacity to analyze correctly the maximum number of patients possible. In this context, methods to improve their workflow can bring better performance to the screening process.

1.3 Thesis organization

The rest of this thesis is divided in four chapters. The second chapter presents related previous works that give a general understanding of the problematic encountered for this research or were used for the elaboration of the proposed method that is presented in details in the third chapter. Then the results of the proposed contribution will be discussed in the fourth chapter to state the advantages and limitations of this work. Finally, a conclusion will close this thesis and bring propositions for future related works. An appendix is also included to provide additional information on some elements mentioned in the course of this dissertation.

CHAPTER 2 PREVIOUS WORK

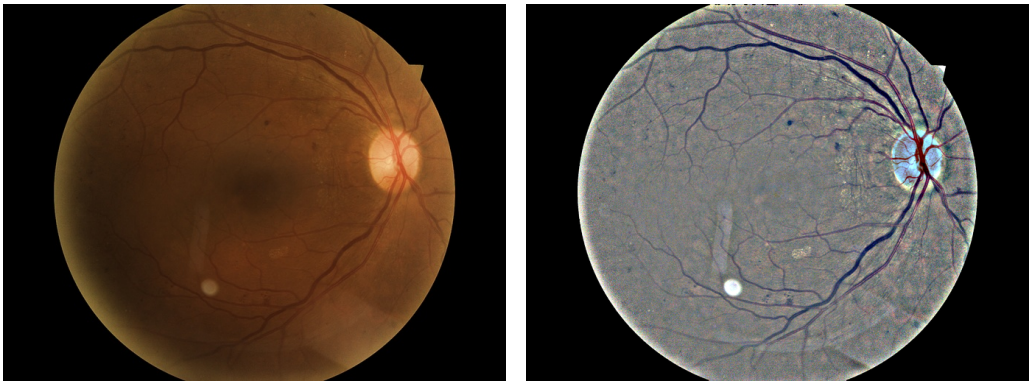
Eye fundus images present a lot of variability in terms of colors which can make the diagnosis more difficult for the graders. The objective of this work is thus to propose a new method to reduce this variability and produce images suited for the grading by a human expert. We are going to present previous works in the two fields of research that focus on similar problems. In the first section, we will consider image enhancement methods targeting the improvement of the image quality whether for human visualization of further CAD system algorithms. In the second section, we will present by-example color manipulation methods that propose solutions to transfer color and atmosphere properties between images. In the last section, we will expose in more details our objectives and hypotheses for this work.

2.1 Eye fundus image enhancement

This manuscript focuses on manipulating eye fundus images to make them easier to analyze by a human expert and to reduce the variability between images. This issue is related to the very active field of image enhancement which has a lot of applications dealing with medical images, and in particular with eye fundus images.

Most previous methods related to the enhancement of eye fundus images focus on the correction of luminosity artifacts such as unequal illumination across the image or low contrast in the image. The current standard for this task is the use of an adaptive unsharp masking [Polesel et al. (2000)] that consists in subtracting to the image a locally smoothed version of itself. One of the main problems of this method is that it makes the fovea disappear as can be seen in Figure 2.1. More recent solutions propose techniques that do not have this problem. The first one uses the Mahalanobis Distance (MD) to classify pixels into background pixels or foreground pixels that are respectively the background of the eye and all the other elements in the image [Foracchia et al. (2005)]. It then corrects the illumination drifts using the hypothesis that the background can be modeled as a white random field. An other method uses a technique called Contrast Limited Adaptive Histogram Equalization (CLAHE) that is a form of Adaptive Histogram Equalization (AHE) [Setiawan et al. (2013)]. AHE is an image enhancement process in which the image is divided into subparts for which a separate histogram equalization is performed. CLAHE is an improvement of this method as it prevents the noise in the image to be enhanced by limiting the contrast enhancement of each AHE.

These methods are generally used as a preprocessing step to other advanced algorithms such as retinal vessels segmentation, since the resulting image with enhanced contrast and equalized illumination are better input for these systems. They are also sometimes used as visualization methods based on the argument that better contrast should help experts make their diagnosis. However, the produced images are in reality not popular among graders as they present unnatural colors, as can be seen in Figure 2.1, that they find harder to interpret. These methods are then not appropriate for the objective of our work.



(a) Source (b) Enhanced

Figure 2.1 Enhancemenet by unsharp masking. While the luminosity and contrast in the image is corrected, the fovea, which is the darker region at the center of the image, is removed by the enhancement method and the color information is deteriorated. (Image courtesy of Dr. Chakor)

2.2 By-example color manipulation

As we do not want to deteriorate the structures in the images since that could create errors in the diagnosis, we consider methods that only modify the colors of the pixels in the image.

By-example color manipulation is the field of research focusing on transferring the colors of a given target image used as a reference to another input image. It is an example-based image processing technique used for two separate applications that are the colorization and the automatic color transfer of an image. Colorization is the process of adding colors to an originally non-colored image whereas automatic color transfer changes the colors of an already colored image. Examples for both these applications on natural images can be found on Figures 2.2 and 2.3. These techniques are commonly used for changing the atmosphere of images in the case of regular photography and only recently has there been some applications to a medical context. Until now, only colorization has been considered for

medical images [Nida and Khan (2016); Popowicz and Smolka (2015); Zhao et al. (2007)] as they tend to be in gray scale. However, retinal image photographies are already in color but no previous work tried to use automatic color transfer on them. As our objective is to design a method that would produce images adapted for visualization by a human expert, we do not consider the colorization methods to be appropriate because using only one color channel would remove some color information in the image that can help understand and better describe its content. This thesis will thus concentrate on the automatic color transfer methods.

(a) Source (b) Target (c) Output

Figure 2.2 By-example colorization example. (a) is the input image to colorize. (b) is the target image with the wanted colors. (c) is the resulting image from the colorization of (a) by (b). (Images from [Arbelot et al. (2016)])

(a) Source (b) Target (c) Output

Figure 2.3 By-example color transfer example. (a) is the input image of which to change the colors. (b) is the target image with the wanted colors. (c) is the resulting image from the color transfer of (b) to (a). (Images from [Arbelot et al. (2016)])

2.2.1 Global color transfer

The precursor work on this field uses global color statistics in the decorrelated $L\alpha\beta$ color space to do a reshaping of the histogram of the source image [Reinhard et al. (2001)]. For each channel, it uses the mean and variance of the corresponding channel of the target image. It relies on the hypothesis that the color distribution on each channel is a Gaussian. While it can produce good results with some sets of images, artifacts tend to be visible with this method as colors not present in either of the two original images appear in the resulting one. To try to solve this problem, other works proposed improvements built upon this initial simple method. One used a multidimensional probability density function to model all channels at once [Pitié et al. (2007)]. More recently, a multi-scale approach proposed to match the histograms of each channel at various scales allowing for a better fitting of the smaller features [Pouli and Reinhard (2011)]. These global methods have overall good results but still present some issues as histogram matching considers the image at the highest level. As such, they still produce in some cases new colors not present in the original images and we have no insurance that similar regions will have similar colors after the transfer. Another concern is that the contours in the image can not be respected while applying the new colors. To solve this issue, a solution proposed to view it as an optimization problem, trying to preserve the color gradient from the source image while applying colors from the reference image [Xiao and Ma (2009)]. Other methods have also approached the subject as an optimal transportation problem [Ferradans et al. (2013)]. But these solutions still do not ensure that similar regions are correctly transferred between each other since more local information should be taken into account.

2.2.2 Local color transfer

Several methods use more local information to try to ensure this constraint is respected. A few of them require user input [Freedman and Kisilev (2010)] and are thus outside of the scope of our work which aims at being automatic. Another method uses a Gaussian Mixture Model to segment the source and target images and transfer colors between regions of similar luminance [Tai et al. (2005)]. A more recent method uses color-based segmentation to transfer similar color palettes using optimal transportation [Frigo et al. (2014)]. These solutions usually provide improvements over previous works but still only rely on first order color information. It means that semantically different regions that would present similar colors can still be incorrectly transferred between one another. Other methods rely on spatial correspondence in the images [HaCohen et al. (2011); Faridul et al. (2013); Hwang et al. (2014)] but are specifically designed to work on images representing the exact same scene

under different conditions or angles such as consecutive frames in a video or pictures of the same scene with different exposures. Thus, they are not well adapted to our problem. Another work proposed to separate the images into their foreground and background before segmenting more specifically the background into more specific parts [Wu et al. (2013)]. After a matching process, the transfer is then executed between regions of similar semantic. While presenting an interesting usage of the semantic content of the images, this method relies on detection techniques adapted for natural images such as face detection and does not apply well to our problem either. Finally, a recent work proposed to use color information beyond the first order with a texture descriptor to determine a local color transfer for each pixel [Arbelot et al. (2016)]. It achieves better results when applied to natural images while not being tied to a specific type of image or content. It also ensures that the transferred color of each pixel from the source image is more influenced by the colors of similar pixels from the target image.

2.2.3 Comparison methods

For the reasons stated previously and because it is the state of the art, we will use the method from [Arbelot et al. (2016)] as the base for our work. Since the method from [Reinhard et al. (2001)] is the seminal work in the field and presents the general characteristics of the other global color transfer methods, we will also use it as a comparison for our method. In the following sections, we describe in more details these methods.

Reinhard et al. (2001)

The first step of this method is to convert the images from the RGB color space into the $L\alpha\beta$ color space, which is a decorrelated color space. This allows the further steps of the method to be applied to each channel of this color space independently. We thus describe the process for a single channel that we note c, the operations for all the channels being the same.

The next step is to compute the mean and standard deviation of the channel c for both the source and the target images. We note μ^{source} and σ^{source} the mean and standard deviation of the source image and μ^{target} and σ^{target} the mean and standard deviation of the target image.

These values are then used to correct the colors of the channel, using the following equation for all the pixel $\boldsymbol{p}$ of the channel c of the source image:

$$\mathcal{T}(\boldsymbol{p}) = \frac{\sigma^{target}}{\sigma^{source}}(c^{source}(\boldsymbol{p}) - \mu^{source}) + \mu^{target} \tag{2.1}$$

After the channels have been computed separately, the final step is to convert the resulting image back to the original RGB color space.

Arbelot et al. (2016)

This method is based on the same idea as the previous one to correct the mean and standard deviation of the image. However, instead of using only global statistics, it introduces a local texture descriptor to modify how each pixel is transferred. As with the previous method, the first step is to convert the image into a decorrelated color space to be able to work on each channel separately, in this case the CIELAB color space. As was done for the previous method, the following steps are described for a single channel c.

The next step is to compute a similarity between the pixels of the images using a texture descriptor. This descriptor is based on a 6-dimensional feature vector for each pixel $\boldsymbol{q}$:

$$\boldsymbol{z}(\boldsymbol{q}) = \begin{bmatrix} L(\boldsymbol{q}) & \frac{\delta L(\boldsymbol{q})}{\delta x} & \frac{\delta L(\boldsymbol{q})}{\delta y} & \frac{\delta^2 L(\boldsymbol{q})}{\delta x^2} & \frac{\delta^2 L(\boldsymbol{q})}{\delta y^2} & \frac{\delta^2 L(\boldsymbol{q})}{\delta xy} \end{bmatrix}^T \tag{2.2}$$

where $L(\boldsymbol{q})$ is the luminance of the pixel $\boldsymbol{q}$. A region descriptor is then defined using the covariance matrix for the feature vectors in each region. Using this region descriptor, a similarity metric is defined to measure the similarity between regions in the images: $\mathcal{D}(\boldsymbol{p}, \boldsymbol{q})$ is the similarity between the regions centered on the pixels $\boldsymbol{p}$ and $\boldsymbol{q}$.

As in the previous method, we then compute the means and standard deviations of the channels. However, instead of using the same values for all the pixels, we compute a specific value for each pixel of the source using the similarity metric. For a given pixel $\boldsymbol{p}$ from the source image, we note these values $\mu^{source}(\boldsymbol{p})$, $\sigma^{source}(\boldsymbol{p})$, $\mu^{target}(\boldsymbol{p})$ and $\sigma^{target}(\boldsymbol{p})$. For each pixel, we obtain these ponderated means and standard deviations using the following equations:

$$\begin{aligned} \mu^{img}(\boldsymbol{p}) &= \frac{1}{W} \sum_{\boldsymbol{q} \in c^{img}} c^{img}(\boldsymbol{q}) \mathcal{D}(\boldsymbol{p}, \boldsymbol{q}^{img}) \\ \sigma^{img}(\boldsymbol{p}) &= \sqrt{\frac{1}{W} \sum_{\boldsymbol{q} \in c^{img}} (c^{img}(\boldsymbol{q}) - \mu^{img}(\boldsymbol{p}))^2 \mathcal{D}(\boldsymbol{p}, \boldsymbol{q}^{img})} \end{aligned} \tag{2.3}$$

Where img is either $source$ or $target$ and W is the normalization factor: $W = \sum_{\boldsymbol{q} \in c^{img}} \mathcal{D}(\boldsymbol{p}, \boldsymbol{q}^{img})$. These values are then used to correct the colors the same way as in the previous method:

$$\mathcal{T}(\boldsymbol{p}) = \frac{\sigma^{target}(\boldsymbol{p})}{\sigma^{source}(\boldsymbol{p})} (c^{source}(\boldsymbol{p}) - \mu^{source}(\boldsymbol{p})) + \mu^{target}(\boldsymbol{p}) \tag{2.4}$$

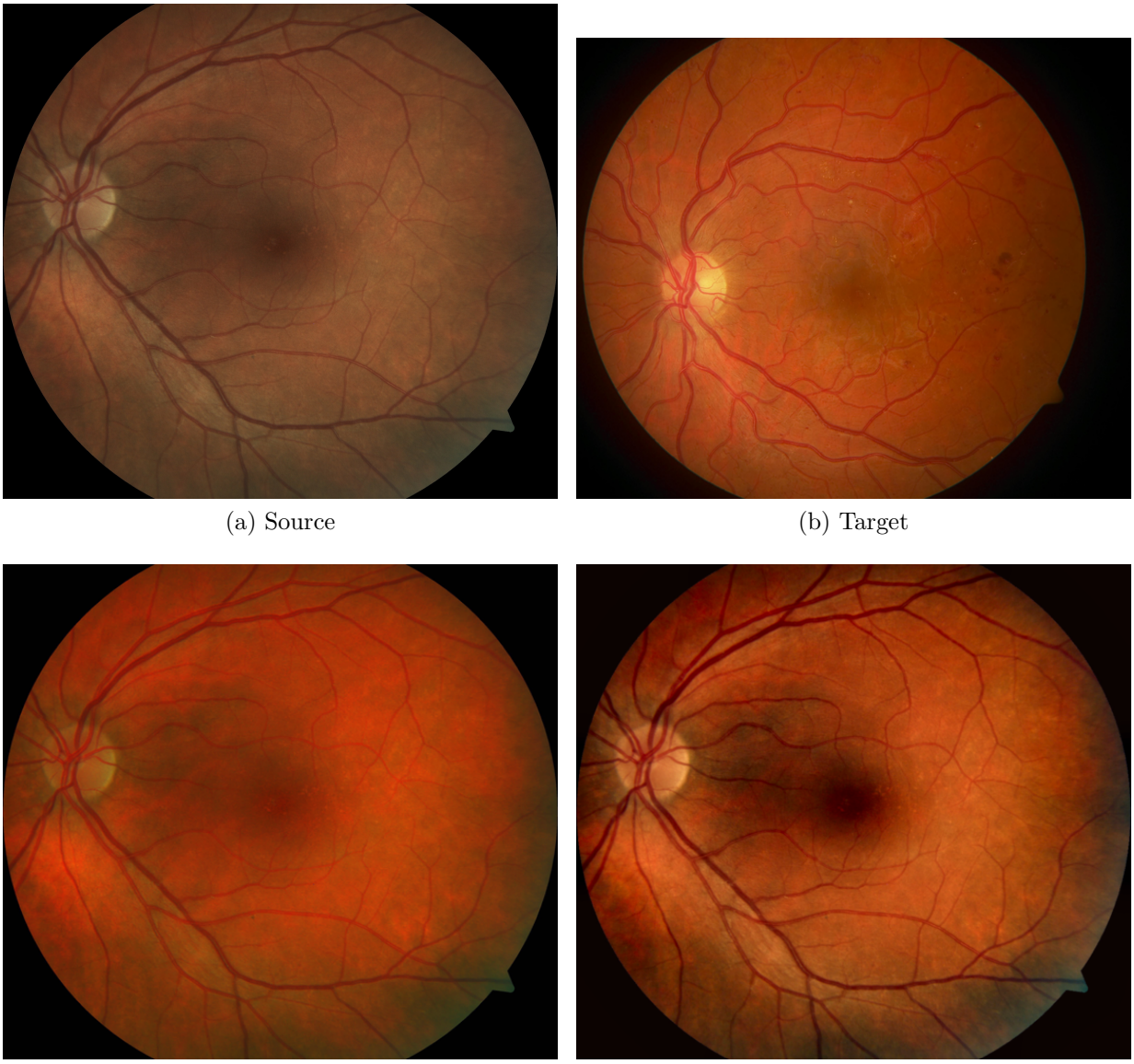
After the channels have been transferred separately, the final step is then to convert the corrected image back to the RGB color space.

As can be seen on Figure 2.4, the method from [Arbelot et al. (2016)] already provides better results than the method from [Reinhard et al. (2001)]. However the transferred image is too dark, especially the vessels, and globally the colors are not close enough from the target image. Some of these problems come from the black pixels outside the camera field of view that influence the final color of each pixel even if the distance used to guide the transfer limits this influence. But we also think that additional information on some anatomical elements such as the vessels would improve the final results by ensuring that their color would match the color of the corresponding elements in the target.

2.3 Objectives and hypotheses

Our main objective with this work is to provide a new automatic visualization method that will help graders to visualize eye fundus images with better precision and in a lesser amount of time. Our main hypothesis is that having all the images in the same color palette of an image of good quality will allow for an easier grading. It is very important to ensure that the resulting image is as close as possible of the target image chosen by the expert for two main reasons. The first is that any difference in the result on any anatomical element would harm the trust in the resulting image from the graders as these elements would not be in the expected color. The second is that the target image represents the preferences of the expert in terms of colors, and thus the closer our result is from this reference, the better it would be for the grader.

More specifically, we aim at providing an automatic color transfer method that is able to distinguish and match between themselves the different anatomic elements in the images. Our hypothesis is that using texture information in the images as well as vascular vessels segmentation to guide the color transfer will allow for a matching of these elements in terms of colors.



(a) Source

(b) Target

(c) Reinhard et al. (2001)

(d) Arbelot et al. (2016)

Figure 2.4 Comparison of some previous works on color transfer. (a) is the original image used as a source. (b) is the target image. (c) is the transferred image using Reinhard and Pouli (2011). (d) is the transferred image using Arbelot et al. (2016). (Images courtesy of Dr. Chakor)

CHAPTER 3 METHODOLOGY

In this chapter we present our proposed method of image enhancement for the visualization of eye fundus images. Section 3.1 presents a general overview of the method and the different steps that compose it. Section 3.2 presents our method to extract the Region Of Interest (ROI), Section 3.3 details the vessels segmentation technique used. Section 3.4 presents the color transfer method in details with Section 3.4.1 describing the similarity metric used to guide the transfer, Section 3.4.2 the color transfer applied using the previously computed information and Section 3.4.3 an optimization used to speed up the process.

Our overall contribution is to study the application and adaptation of a color transfer technique in the context of eye fundus visualization towards the improvement of manual screening by experts. The color manipulation part of our methodology is strongly based on a consumer photo color manipulation technique that we extend and adapt to the context of eye fundus images. In Section 3.4.1, we propose a new similarity metric to take into account the vascular vessels segmentation in order to improve the color matching of elements in the images. In Section 3.4.3, we present how we adapt the quantization to take into consideration the semantic information of the images to reach a balance between performance and color precision.

3.1 Method overview

Our proposed method consists of multiple successive steps that operate on the source image of which we want to change the colors and the target image whose color we want to use. The final output is the source image transferred to the colors of the target. The first steps consist in segmenting the Region Of Interest (ROI) and the vascular network. The following steps are based on the chosen color transfer method [Arbelot et al. (2016)] which we adapt to our context. Using the previously computed information and the images, we compute the similarity between the pixels. The additional information that we introduce allow us to better match the anatomical structures in the images, and thus to provide a better guide for the color transfer. Using this similarity metric, the ROI and the images, we apply a quantified color transfer to obtain the output of our system. To speed up the process, an intermediary step of quantization is added before the color transfer. The complete process can be observed in Figure 3.1.

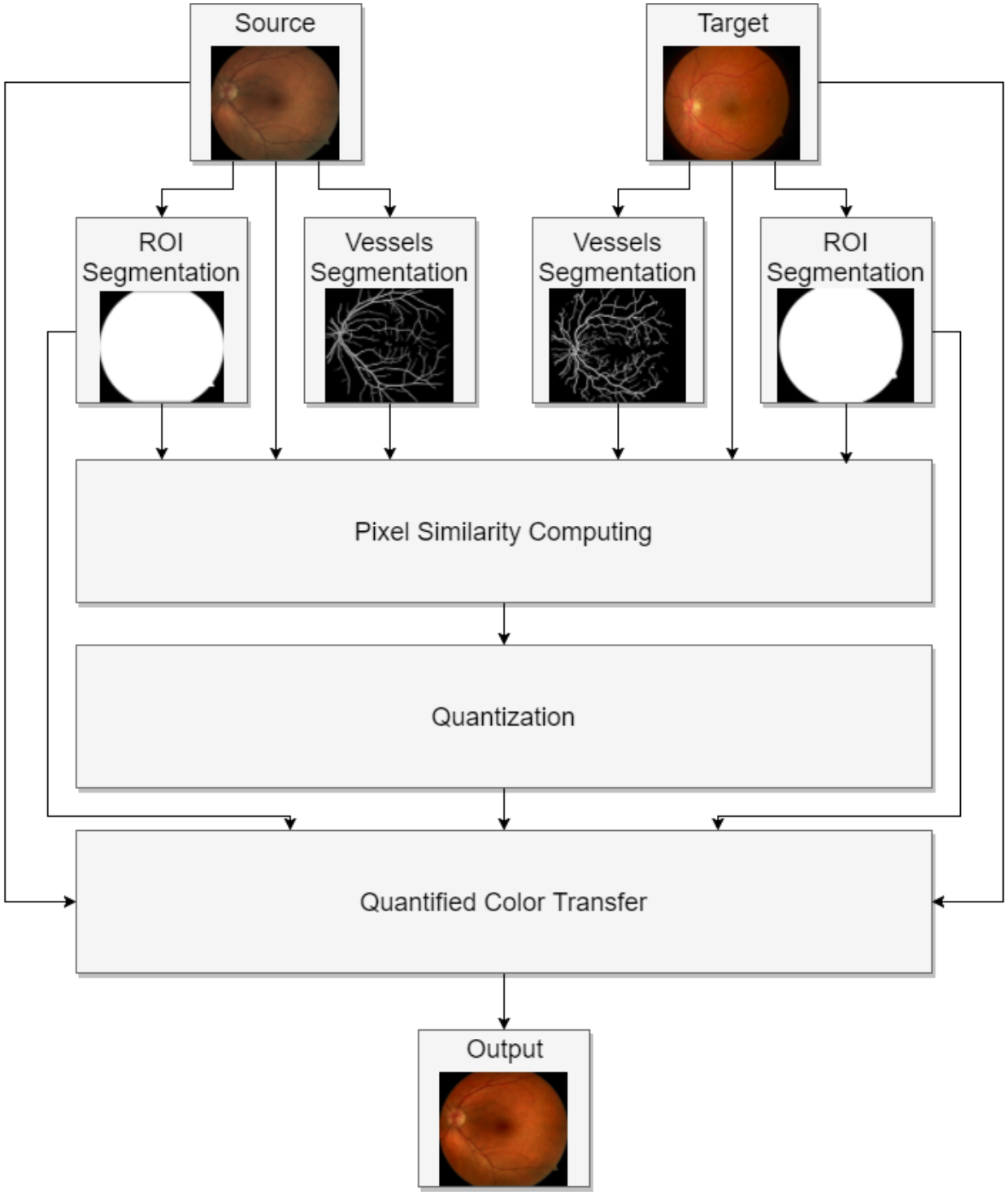
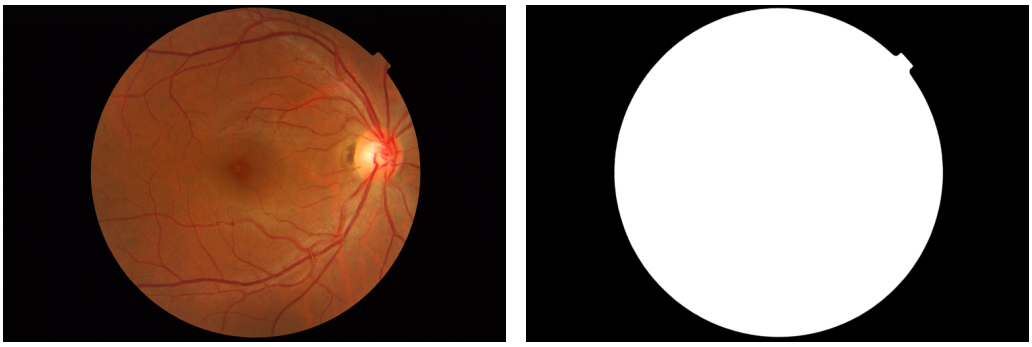


Figure 3.1 Pipeline of the proposed method.

3.2 Segmentation of the region of interest

The first problem that we have to tackle is the shape of our eye fundus photography images. In these images, there are black pixels outside the camera field of view that we do not want to be considered since we are only interested in the actual content, the photography of the fundus. We define the fundus pixels as our ROI. This ROI can vary from a camera to another and is not necessarily provided by the camera. Therefore we need an automatic method to get a segmentation mask.

Our method for this task has four main steps. The first step is to sum all the channels into one so that all the information of the original image is then contained into a single intensity channel. We then denoise this channel by applying a median blur with a 3-by-3 kernel. We then threshold the channel at 10% of its maximum value since the unwanted pixels have a value close to 0. To remove background noise, we apply successively the morphologic operations of opening and closing with a 3-by-3 kernel. Finally, to ensure that our mask only contains pixels from our ROI, we apply an operation of erosion with a 11-by-11 kernel which in particular allows us to remove peripheral pixels that are sometimes blurred by the camera. We chose the parameters so that the method works without having to change the values between our images and without getting any pixel outside of our ROI. An example of a mask produced by this method can be found in Figure 3.2.



(a) Source (b) ROI Mask

Figure 3.2 Sample result of our ROI segmentation method. (a) is the original image used as a source. (b) is the output mask of our method. (Images courtesy of Dr. Chakor)

3.3 Segmentation of the vascular network

The second step of our method is the automatic segmentation of the vascular network as we want to use this information to improve the matching of the anatomic elements of the

eye fundus images. For this task, we use the state-of-the-art method using Multi-Scale Line Detection (MSLD) [Wang et al. (2009); Nguyen et al. (2013)] with the preprocessing steps, the adaptive thresholding and the post-processing step described by a recent work on the subject [Christodoulidis et al. (2016)] and the parameters values it proposed. In this section, all our operations only consider the green channel of the image as it is the channel in which the contrast between the vessels and the background is the highest.

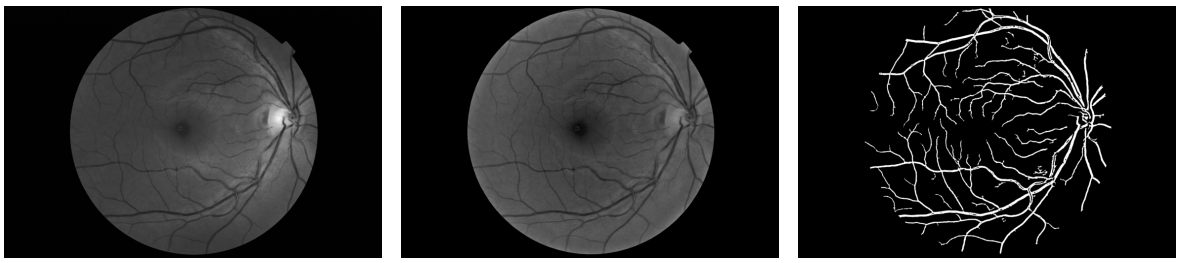
The preprocessing of the image for the segmentation of the vascular network consists in two steps. The first one [Foracchia et al. (2005)] uses a method based on the Mahalanobis Distance (MD) to correct the non-uniform contrast and luminosity in the image while the second one is a denoising operation using a Discrete Wavelet Transform (DWT) [Donoho (1995)]. It is based on the idea that the images can be represented as the sum of its foreground, which contains only the high frequency elements such as the vascular structures, the lesions and the optic disc, and its background which contains the fundus image without these elements. The background image is statistically modeled as a white random field $\mathcal{N}(\mu_b, \sigma_b)$ where μ_b represents the ideally uniform luminosity value and σ_b the natural variability of the fundus. This representation is then used to extract the background pixels using the MD:

$$d_M = \frac{I(x,y) - \hat{\mu}_N}{\hat{\sigma}_N} \tag{3.1}$$

which measures the distance between a value and a distribution. The method then computes for each pixel of the background a local mean and standard deviation which represent estimations of the local luminosity and contrast drifts respectively. These local estimations are finally used to normalize the luminosity and contrast across the image.

The second step of the preprocessing removes the background noise with a filtering technique using DWT [Donoho (1995)]. The specific DWT implementation is the Dual-Tree Complex Wavelet Transform [Kingsbury (1998)] which is commonly used in retinal images as it has better directional selectivity than usual DWT, is multi-scale, and approximates shift invariance. An example of result of the complete preprocessing can be seen in Figure 3.3b.

The segmentation of the vascular network is then obtained by using the MSLD method [Nguyen et al. (2013)]. The key idea of this method is to use a line detector at multiple scales and orientations to detect the vessels. For each pixel, the highest value among the responses to these line detectors is kept. Since the vessels in the images are the elements which present a clear orientation, pixels from the vascular structure will have the highest responses to the line detectors. Thus we obtain a segmentation of the vascular network by thresholding the values previously computed, creating a binary map where pixels with a value higher than the



(a) Source green channel (b) Preprocessed green channel (c) Vessel network segmentation

Figure 3.3 Results of the preprocessing and of the vascular network segmentation. (a) is the original green channel of the input image (b) is the same channel after the preprocessing (c) is the vascular network segmentation of the input image

given threshold are white. By fitting a Gaussian to the histogram of the MSLD response, a threshold value T that separates the fundus from the vessels is obtained:

$$T = \lfloor \mu_{Gaussian} \rfloor + 1.55 \lfloor \sigma_{Gaussian} \rfloor \tag{3.2}$$

After this step, some background noise remains. In order to remove it, we use the post-processing step described by Arbelot et al. (2016). It is a morphological cleaning that removes the smaller connected components that are not elongated enough and not very solid. The solidity is the ratio of the area covered by the connected component over its convex hull area. This corresponds to removing the elements whose shape and size least correspond to the vascular network. An example of vascular network segmentation can be observed in Figure 3.3c.

3.4 Content-based color transfer

In this section, we describe the steps to apply the color transfer between the target and the input images using the previously computed information along with the images. The target image that we use in the figures illustrating the following sections has been chosen by an ophthalmologist.

3.4.1 Pixel similarity computing

The first step is to compute the similarity between the pixels. In the following section, we present the similarity metric introduced by Arbelot et al. (2016) along with the changes we propose.

Descriptor similarity

Arbelot et al. (2016)'s key idea is to describe each pixel $\boldsymbol{q}$ by a 6-dimensional feature vector:

$$\boldsymbol{z}(\boldsymbol{q}) = \begin{bmatrix} L(\boldsymbol{q}) & \frac{\delta L(\boldsymbol{q})}{\delta x} & \frac{\delta L(\boldsymbol{q})}{\delta y} & \frac{\delta^2 L(\boldsymbol{q})}{\delta x^2} & \frac{\delta^2 L(\boldsymbol{q})}{\delta y^2} & \frac{\delta^2 L(\boldsymbol{q})}{\delta xy} \end{bmatrix}^T \tag{3.3}$$

where $L(\boldsymbol{q})$ is the luminance of pixel $\boldsymbol{q}$. Using this feature vector, square regions of size $(2r+1) \times (2r+1)$ centered on a given pixel $\boldsymbol{p}$ that we note $N_r^{\boldsymbol{p}}$ are represented as the 6-by-6 covariance matrix:

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{p}) &= \frac{1}{W} \sum_{\boldsymbol{q} \in N^{\boldsymbol{p}}} (\boldsymbol{z}(\boldsymbol{q}) - \boldsymbol{\mu})(\boldsymbol{z}(\boldsymbol{q}) - \boldsymbol{\mu})^T w(\boldsymbol{p}, \boldsymbol{q}) \\ w(\boldsymbol{p}, \boldsymbol{q}) &= exp(\frac{-9\|\boldsymbol{p} - \boldsymbol{q}\|^2}{2r^2}) \end{aligned} \tag{3.4}$$

where $\boldsymbol{\mu}$ is the vector containing the mean of each feature inside $N^{\boldsymbol{p}}$, $w(\boldsymbol{p}, \boldsymbol{q})$ is a weighting function that ensure that closer pixels have a greater influence on the result, thus enabling a smoother transition of the descriptor from pixel to pixel and $W = \sum_{\boldsymbol{q} \in N^{\boldsymbol{p}}} w(\boldsymbol{p}, \boldsymbol{q})$ is the normalization factor. A visualization of the pixels involved in this equation for $r = 1$ can be observed in Figure 3.4.

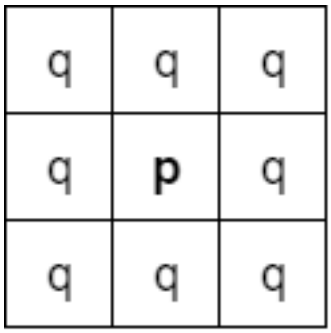


Figure 3.4 Square region visualization for $r = 1$.

This region descriptor is the base for the texture descriptor. However, region covariance matrices only describe second-order statistics, and computing distances between two of them is expensive. For these reasons, the descriptor is further modified following a previously proposed solution to this problem [Karacan et al. (2013)]. This solution uses the Cholesky decomposition $\boldsymbol{C} = \boldsymbol{L}\boldsymbol{L}^T$ where $\boldsymbol{L}$ is a lower triangular matrix. The descriptor is finally represented as a collection of vectors that can more easily be compared and to which is

added some first-order information:

$$\boldsymbol{S} = (\boldsymbol{L}^1 \cdots \boldsymbol{L}^6 \boldsymbol{\mu}) \tag{3.5}$$

where $\boldsymbol{L}^i$ is the i^th column of $\boldsymbol{L}$. Due to the descriptor describing the region around a given pixel, edges tend to be blurred with it. To ensure that the edges between the regions are sharp and only localized on edges present on the original image, the method adds two additional steps before computing the similarity between pixels. The first step is a multi-scale gradient descent that sharpens the edges. The second step is an unnormalized bilateral filter [Aubry et al. (2014)] modified to use the luminance image as the guide. This step makes the edges sharper with finer details and reduces the variability inside the regions.

Finally, the similarity between two pixel using the obtained descriptor map is introduced:

$$\mathcal{D}_{\sigma_d}(\boldsymbol{p}, \boldsymbol{q}) = exp\left(\frac{-\|\boldsymbol{S}(\boldsymbol{p}) - \boldsymbol{S}(\boldsymbol{q})\|^2}{2\sigma_d^2}\right) \tag{3.6}$$

where $\boldsymbol{S}(\boldsymbol{p})$ and $\boldsymbol{S}(\boldsymbol{q})$ are the descriptor of pixels $\boldsymbol{p}$ and $\boldsymbol{q}$ respectively and σ_d is the standard deviation that controls the impact of a difference in descriptor on the similarity.

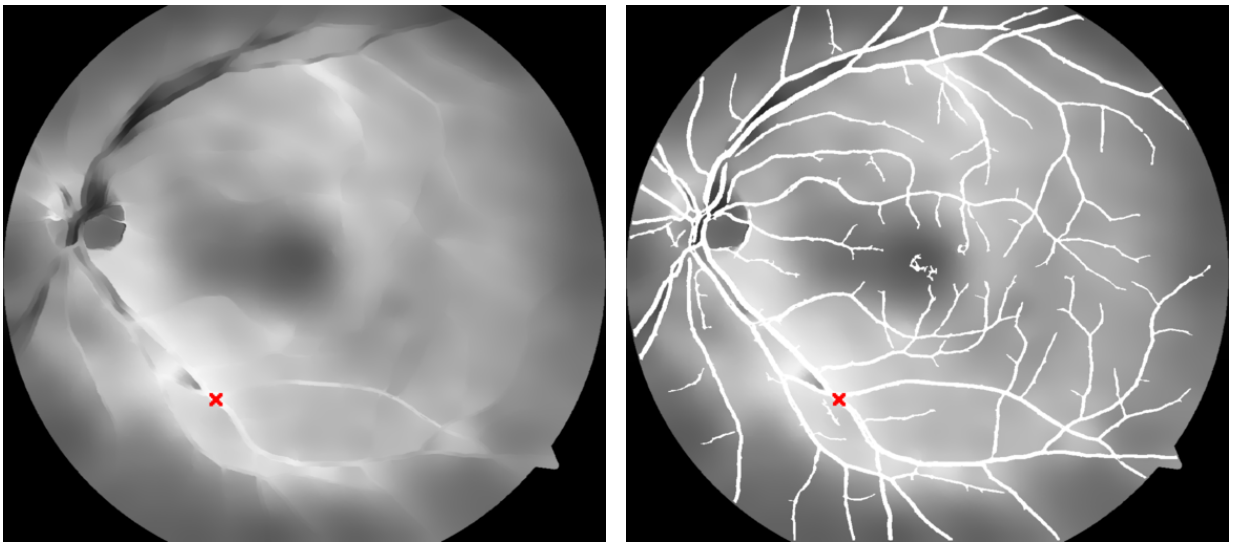
Vessels segmentation factor

We propose to modify Equation (3.6) so that pixels belonging to the same semantics are considered closer to each other than to other pixels. For this purpose, we use the segmentation of the vascular network to adapt the formula of the similarity depending on whether the two compared pixels belong to the segmentation or not. If both pixels belong to the segmentation, we multiply the similarity value between them by a factor whose value depends on a parameter f. If one or the other is not a vessel pixel, then the similarity value is left as is. If we note $s(\boldsymbol{p})$ the semantic value of the pixel $\boldsymbol{p}$ with $s(\boldsymbol{p}) = 1$ for a pixel in the segmentation and $s(\boldsymbol{p}) = 0$ for a pixel outside of it, then we have our new similarity metric:

$$\tilde{\mathcal{D}}_{\sigma_d}(\boldsymbol{p}, \boldsymbol{q}) = (1 + f \times s(\boldsymbol{p}) \times s(\boldsymbol{q}))\mathcal{D}_{\sigma_d}(\boldsymbol{p}, \boldsymbol{q}) \tag{3.7}$$

This modification allows for a smooth integration of the semantic information on top of the texture descriptor. This allows for a better matching of the vessels between each other while still considering the variability that can occur within them. Still relying on the texture descriptor is necessary to correct some segmentation errors. Indeed, segmentation algorithms often miss some vessels and sometimes also include other elements, in particular some lesions

that closely resemble small vascular vessels but which generally have very different textural properties. Similarity maps for a pixel in the vessel network with and without the vessels segmentation factor can be observed in Figure 3.5.



(a) Without the vessels segmentation factor (b) With the vessels segmentation factor

Figure 3.5 Similarity maps for a pixel in the vessel network. The pixel considered for these similarity maps is represented by the red crosses. Pixels in the vessel network are considered more similar when using our vessel segmentation factor. (Images courtesy of Dr. Chakor)

3.4.2 Quantified color transfer

As in Arbelot et al. (2016), we combine the global color transfer model of Reinhard et al. (2001) with the proposed similarity metric in order to compute an adaptive transfer map for each pixel. Therefore, instead of using global uniform statistics, the method computes a weighted mean and standard deviation for each channel c as follow:

$$\begin{aligned} \mu^{img}(\boldsymbol{p}) &= \frac{1}{W}\sum_{\boldsymbol{q}} c^{img}(\boldsymbol{q})\tilde{\mathcal{D}}_{\sigma_d}(\boldsymbol{p}^{in},\boldsymbol{q}^{img}) \\ std^{img}(\boldsymbol{p}) &= \sqrt{\frac{1}{W}\sum_{\boldsymbol{q}}(c^{img}(\boldsymbol{q})-\mu^{img}(\boldsymbol{p}))^2\tilde{\mathcal{D}}_{\sigma_d}(\boldsymbol{p}^{in},\boldsymbol{q}^{img})} \end{aligned} \tag{3.8}$$

With $img \in in, ref$ where in refers to the original input image, ref to the target image and $\boldsymbol{q}$ iterates over the ROI of img. $W = \sum_q \tilde{\mathcal{D}}_{\sigma_d}(\boldsymbol{p}^{in},\boldsymbol{q}^{img})$ is the normalization factor. Each

pixel is then modified with the transfer function as originally defined:

$$\mathcal{T}_{\sigma_d}(\boldsymbol{p}) = \frac{std^{ref}(\boldsymbol{p})}{std^{in}(\boldsymbol{p})}\left(c^{in}(\boldsymbol{p}) - \mu^{in}(\boldsymbol{p})\right) + \mu^{ref}(\boldsymbol{p}) \tag{3.9}$$

Color transfer is usually applied in a decorrelated color space so that each channel can be modified independently. Unlike [Reinhard et al. (2001)] who use the $L\alpha\beta$ color space, we use the CIELab color space with the illuminant E because recent research demonstrated that it is the most adequate choice for color transfer applications [Reinhard and Pouli (2011)].

3.4.3 Computing time optimization by quantization

The objective of our work being to improve the screening process for retinal diseases, we want to ensure that the proposed method runs fast enough to be considered into an image processing pipeline in this context. In order to speed up the color transfer, we propose a quantization based on the similarity metric and the semantic information to group pixels that are then transferred the same way. As in Arbelot et al. (2016), we use a threshold to determine if two pixels of the input image are close enough in term of descriptors to be transferred using the same transformation. However, we propose to consider the semantic information to avoid fundus and vessels pixels to be grouped together. With our method, two pixels that are not part of the same ensemble as defined by the vascular network segmentation cannot be grouped in a single transformation. The quantization using a threshold allows for a faster computation as fewer similarity maps, weighted means and standard deviations have to be computed. Our use of the semantic information ensures that the transformation of the vessels and the background pixels stay separated even in case where they could have been close in the descriptor space. This reduces the risk of wrong color matching between anatomical elements appearing because of the quantization.

CHAPTER 4 RESULTS AND DISCUSSION

In this chapter we discuss the results of our proposed method to verify that we reach our initial objective of transferring the color palette from a given target to an input image. More specifically, we want the colors of the elements in our results to be the closest possible to these same elements in the target image. Completing this objective would produce the intended reduced color variability among images and contribute for the expected easier grading.

In the first section we present some results of our method to demonstrate its effect on reducing the variability among retinal fundus images. We compare it to other color transfer methods to show the improvements brought by our method. We then propose and implement an experimental protocol to validate our hypothesis on the effect of this reduction on the work of the graders. We conclude with a discussion of our method and an assessment of its limitations.

4.1 Comparison with previous works

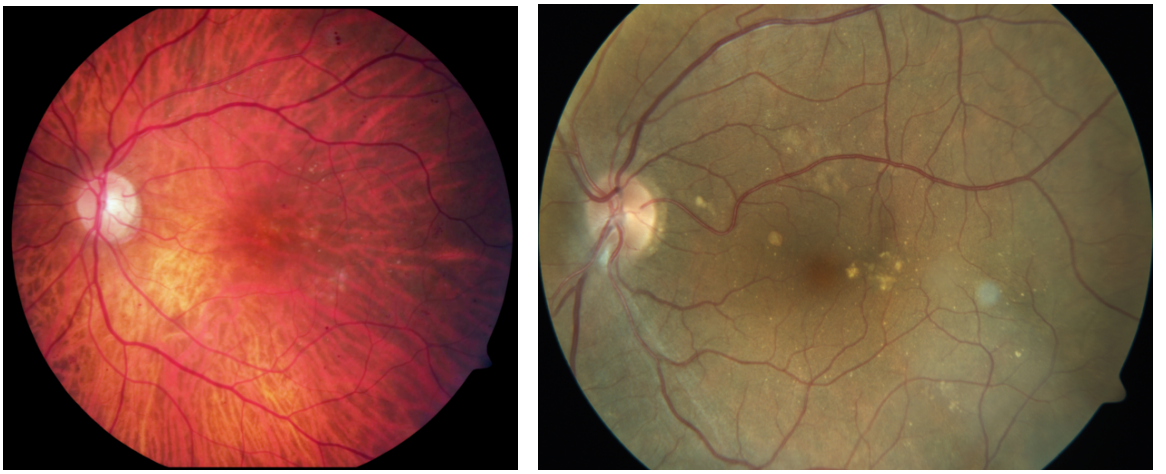
4.1.1 Images comparisons

We compare our method to two previous works. The first one [Reinhard et al. (2001)] serves as a baseline as it is the seminal work in the field. The second one [Arbelot et al. (2016)] being the method on which our approach is inspired, we use it to observe the effects of the extensions to fundus images.

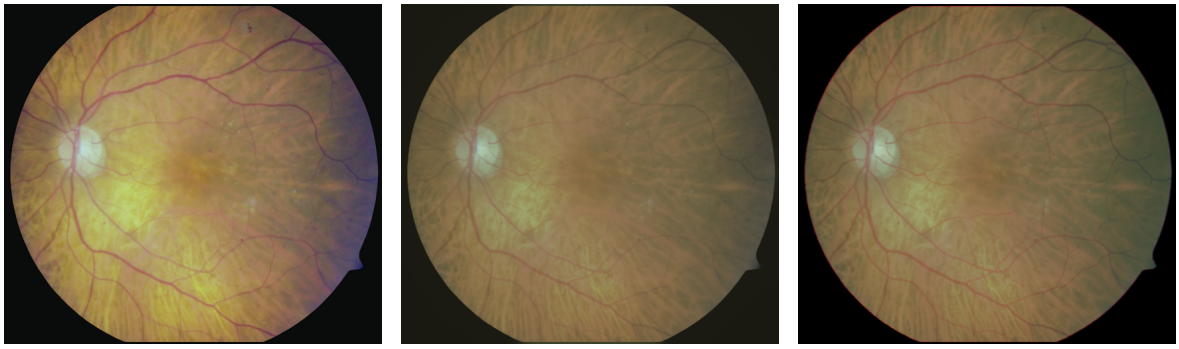
Comparisons with the original methods

In this section we compare our method with the previous works without any modification. In Figure 4.1 and Figure 4.2 we present the results of all the methods for two pairs of images. In Figure 4.1 we use a target with a green tone and red vessels to observe how the methods respect the color of the vessels while transposing the general tone of the image. For these results, we used the original algorithms without any modification and we can see that the method from Reinhard et al. (2001) produces colors that both feel unnatural and were not in the target image. The method from Arbelot et al. (2016) also has some flaws: while it does not produce as visible artifacts as the previous work, it still presents some problems that are more visible in Figure 4.2. Overall, the image is darker than expected, especially for the vessels. While it can be argued that the contrast is better than with our method, we do not find this result satisfying as our objective is to have the final image to use the colors

from the target image.



(a) Source (b) Target



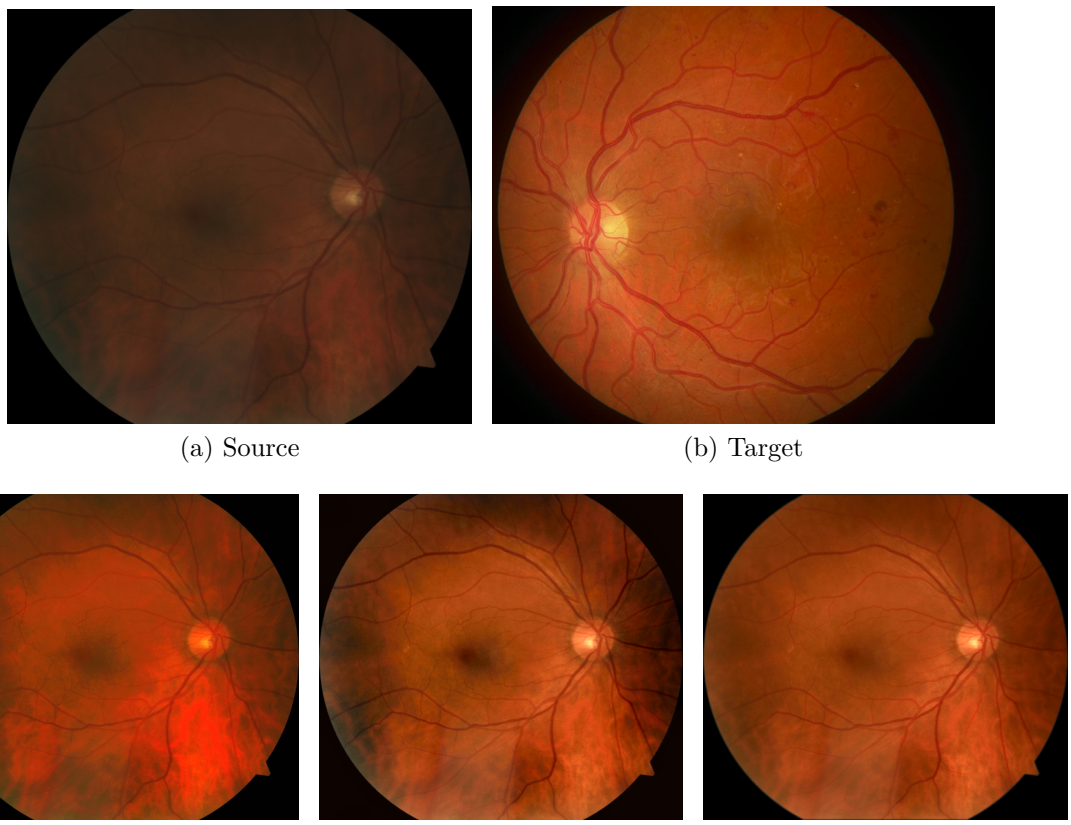
(c) Reinhard et al. (2001) (d) Arbelot et al. (2016) (e) Our result

Figure 4.1 Comparison with the original methods. (c) Reinhard et al. (2001) produces colors that are not present in either of the original images, (d) Arbelot et al. (2016) produces a less color artifacts but with a darker tone than expected (e) our method produces colors closer to the target. (Images courtesy of Dr. Chakor)

The globally darker colors in the image can be explained by the fact that the black pixels outside the camera field of view are taken into account for the transfer since the previous works consider these pixels as regular content. Arbelot et al. (2016) reduces this problem with the texture descriptor that makes the black pixels from the border contribute less to the resulting colors of the pixels from the eye.

Comparisons with methods including ROI segmentation

In this section, we show the effect of the proposed ROI segmentation on the previous methods and then compare them with our method to show the effect of the vascular network



(a) Source (b) Target

(c) Reinhard et al. (2001) (d) Arbelot et al. (2016) (e) Our result

Figure 4.2 Comparison with the original methods. (c) Reinhard et al. (2001) produces colors that are not present in either of the original images, (d) Arbelot et al. (2016) produces a less color artifacts but with a darker tone than expected (e) our method produces colors closer to the target. (Images courtesy of Dr. Chakor)

segmentation.

We modify both of the previous works to use the proposed ROI segmentation and show the produced images in Figure 4.3 and Figure 4.4. This change has a noticeable effect on the result of Reinhard et al. (2001) which produces more correct colors than previously, with Figure 4.3c being the clearest example. The same observation can be made for Arbelot et al. (2016), especially when considering Figure 4.4d.

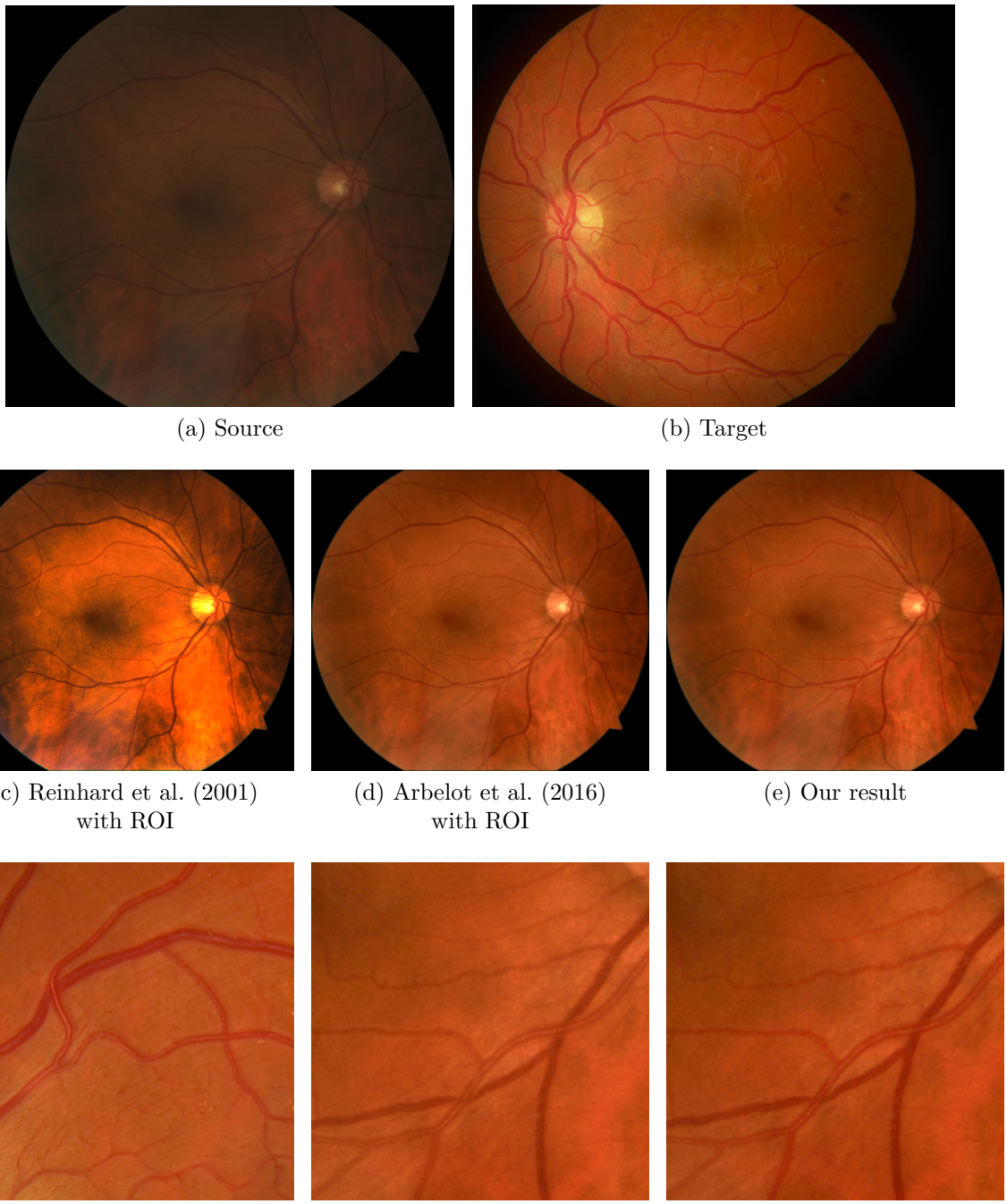
However, in Figure 4.4c we can see that Reinhard et al. (2001)'s result does not have the colors we expect with mostly ones that were not present in the reference image. Moreover, individual elements such as the vessels do not have matching colors. This second problem is present with Arbelot et al. (2016) as well, in particular in Figure 4.4d the vessels are darker than expected.

(a) Source (b) Target

(c) Reinhard et al. (2001) with ROI (d) Arbelot et al. (2016) with ROI (e) Our result

(f) Target (g) Arbelot et al. (2016) with ROI (h) Our result

Figure 4.3 Comparison with the previous works using the ROI segmentation. (c) Reinhard et al. (2001) with the ROI segmentation produces improved results, (d) Arbelot et al. (2016) with the ROI segmentation has tone closer to the target, (e) our method improves the colors of the vessels. (g) with Arbelot et al. (2016) with the ROI segmentation the vessels are too clear, (h) with our approach they are darker and thus closer to those of the target. (Images courtesy of Dr. Chakor)



(a) Source (b) Target

(c) Reinhard et al. (2001) with ROI (d) Arbelot et al. (2016) with ROI (e) Our result

(f) Target (g) Arbelot et al. (2016) with ROI (h) Our result

Figure 4.4 Comparison with the previous works using the ROI segmentation. (c) Reinhard et al. (2001) with the ROI segmentation produces improved results, (d) Arbelot et al. (2016) with the ROI segmentation has tone closer to the target, (e) our method improves the colors of the vessels. (g) with Arbelot et al. (2016) with the ROI segmentation the vessels are too dark and black, (h) with our approach they are redder and thus closer to those of the target. (Images courtesy of Dr. Chakor)

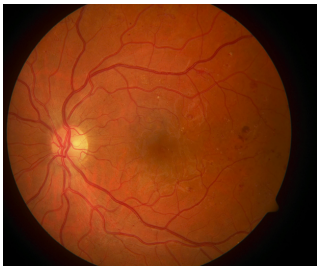
As can be seen in Figure 4.3 and Figure 4.4, especially on the close up of the vessels, our approach of using the vascular network segmentation makes the color of the vessels to be closer to those of the target image. Some other results of our method can be found on Figure 4.5, where it can be observed that all images use the same color palette. We thus think that the objective of reducing the variability among eye fundus images is successfully met.

4.1.2 Histogram comparisons

We now consider the histograms of the results previously presented. In Figure 4.6, we show the normalized histograms of the source, target and our result of Figure 4.3. Also, in Figure 4.7, we present the histograms for the images presented in Figure 4.5. With these we can see that our method reshapes the histograms to make them closer to those of the target. We further demonstrate this observation using the Hellinger distance which measures the distance between two distributions. For two given distributions p and q defined on the same domain X, it can be defined using the Bhattacharyya coefficient as follows [Cha (2007)]:

$$\begin{aligned} D_H(p,q) &= 2\sqrt{1-BC} \\ BC &= \sum_{x \in X} \sqrt{p(x)q(x)} \end{aligned} \tag{4.1}$$

In our case, we measure this distance on each channel separately and use the mean of the distance between the channel as our distance. As is reported in the figures, the distances between the target and our result are consistently smaller than which of the source and the target. However, the final shape is not exactly the same which would be an undesired result. Indeed, the amount of pixels for each element are not the same among images, and thus the distributions should not be the same. In particular, the lesions are not necessarily present in the images and thus the corresponding pixel values from the target should not be matched in our output. Moreover, our method computes a different transformation for the pixels of the source image using local information and thus the color matching is not made at the global level that the histograms represent. Finally, color histograms do not represent how the values of the different channels are combined and thus perfectly matching histograms would not necessarily mean that the images have the same colors.



Target

Source 1

Source 2

Source 3

Source 4

Result 1

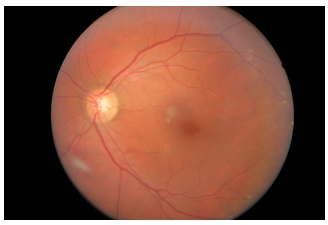
Result 2

Result 3

Result 4

Source 5

Source 6



Source 7

Source 8

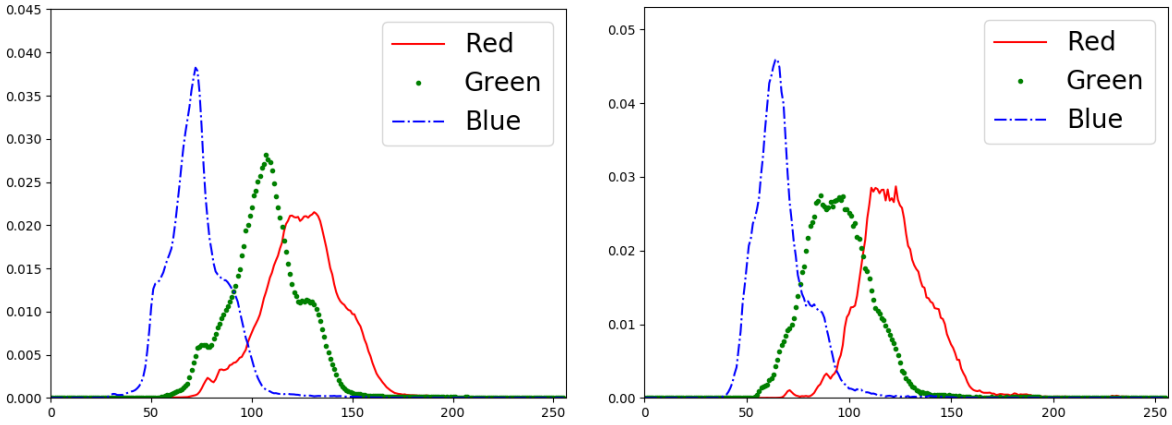
Result 5

Result 6

Result 7

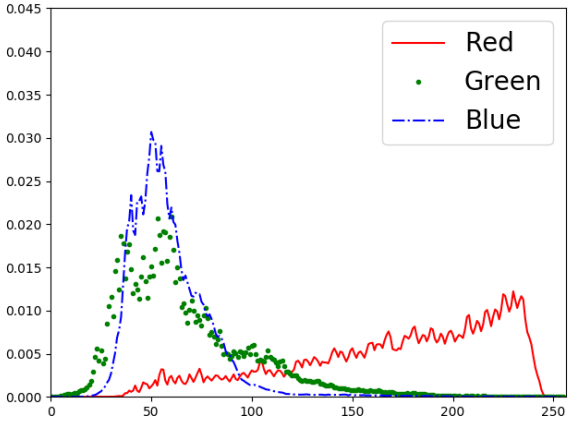
Result 8

Figure 4.5 Results of our proposed method on additional images. (Images courtesy of Dr. Chakor)

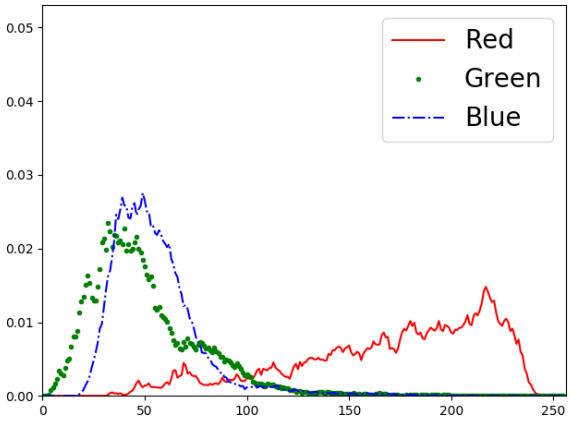


Target

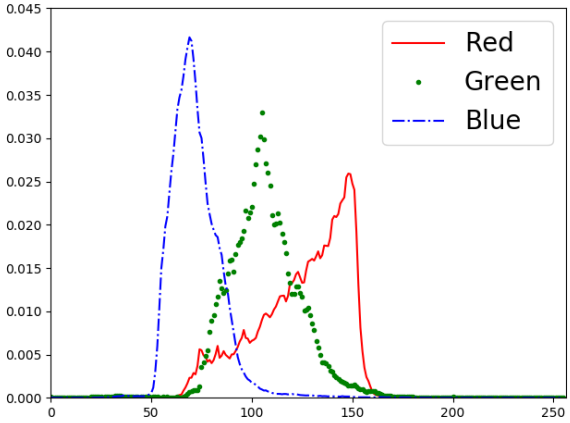
Target (Vessels only)



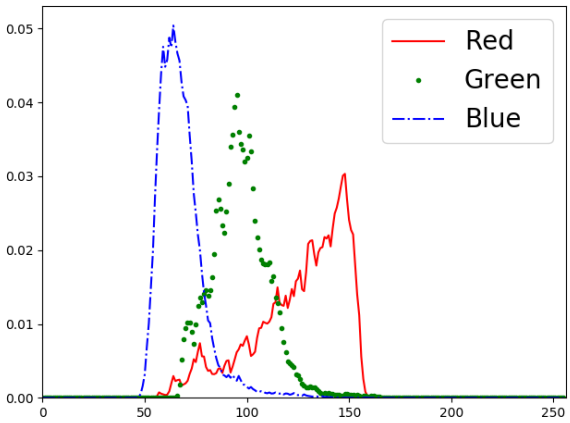
Source
d = 0.55



Source (Vessels only)
d = 0.60

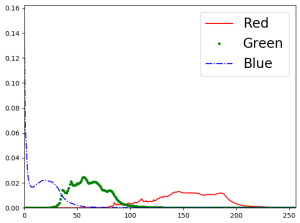


Our result
d = 0.17

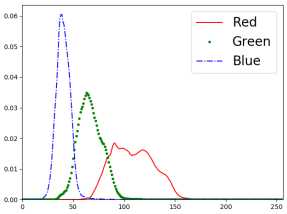


Our result (Vessels only)
d = 0.21

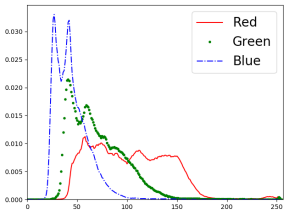
Figure 4.6 Comparison of the histograms of the results from Figure 4.3. The shape of the histogram of the output is closer to which of the target. Below each label is also given the distance to the target in terms of distributions which confirm that the histograms are closer to the histogram of the target after our method is applied.



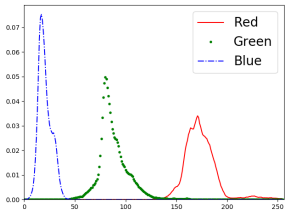
Target



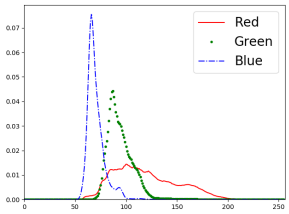
Source 1
d = 0.53



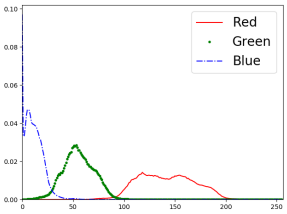
Source 2
d = 0.47



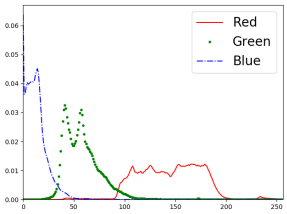
Source 3
d = 0.51



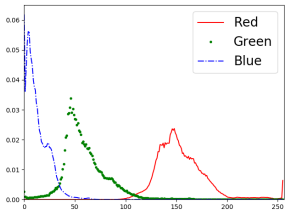
Source 4
d = 0.69



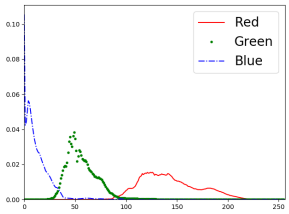
Result 1
d = 0.22



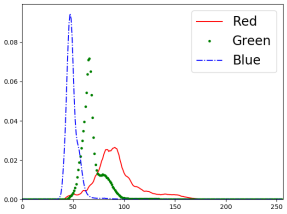
Result 2
d = 0.19



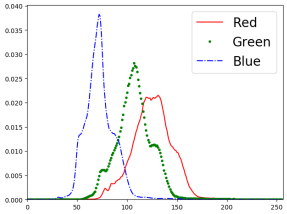
Result 3
d = 0.22



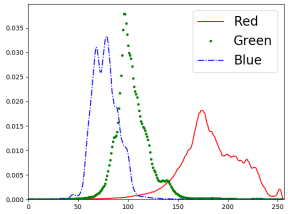
Result 4
d = 0.19



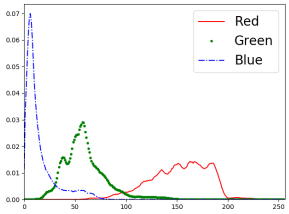
Source 5
d = 0.64



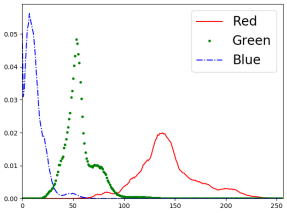
Source 6
d = 0.69



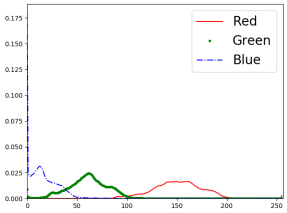
Source 7
d = 0.66



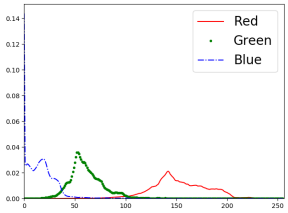
Source 8
d = 0.21



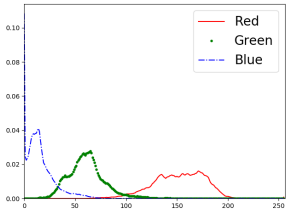
Result 5
d = 0.24



Result 6
d = 0.15



Result 7
d = 0.13



Result 8
d = 0.15

Figure 4.7 Histograms of images from Figure 4.5. Below each label is also given the distance to the target in terms of distributions which confirm that the histograms are closer to the histogram of the target after our method is applied.

4.2 User study

Results in Section 4.1 show that the objective of reducing the variability among images has been met. In order to verify our hypothesis that this can help the graders to make their diagnosis more easily, we designed a user study to estimate the effect of our image enhancement on their work. The objective here is to see how the speed and the precision of the grading are affected positively or negatively by our method. This project, and thus this experiment, is part of a larger project that has been approved by the ethical committee of École Polytechnique de Montréal.

4.2.1 Experimental protocol

We describe here the experimental protocol, starting with some short thoughts that explain our choices. The main idea is to present images modified by our algorithm to users as well as original images to measure the difference between the two cases. Since we believe that having all the images in a single good color palette can make it easier for the graders to give their diagnosis, we decide to present the original and modified images in separate series to be able to measure the ensuing differences in performance between them. Also, we make sure to only show a user a given image once, in either of the two forms that we have so that any change in speed or accuracy cannot be attributed to them remembering their previous diagnosis of the same image. Since a given user won't see both the original and modified versions of an image, we require two different users that will each see the alternative versions of the images seen by the other. If we only depend upon a single user, we would have no way to verify whether an observed increase of speed or accuracy is due to our method or to the images of one of the series being easier than those of the other series.

We now briefly present the protocol, a complete version can be found in the Annexe A and a schematic visualization of it is presented in Figure 4.8. In this experience, we will have two types of measures. The first one is average time spent by the users for the grading of an image. The second one is the error rate in their diagnosis. Our data is comprised of 300 randomly selected images from the image database of Dr. Chakor that we randomly divide into three sets of 100 images. We use our color transfer method on the images from two of those groups of images and get our final sets:

- set_0 which contains 100 non-modified images.
- $set_1^{original}$ and $set_1^{enhanced}$ which contains 100 other images, the first one in their original version, the second one enhanced with our method.

- $set_2^{original}$ and $set_2^{enhanced}$ which contains 100 other images, the first one in their original version, the second one enhanced with our method.

These sets have different purposes. We use set_0 to obtain reference values for the time and error rate of the users in normal conditions. The four remaining sets are used to measure the effect of our method on the work of the users, compared to the references defined with set_0.

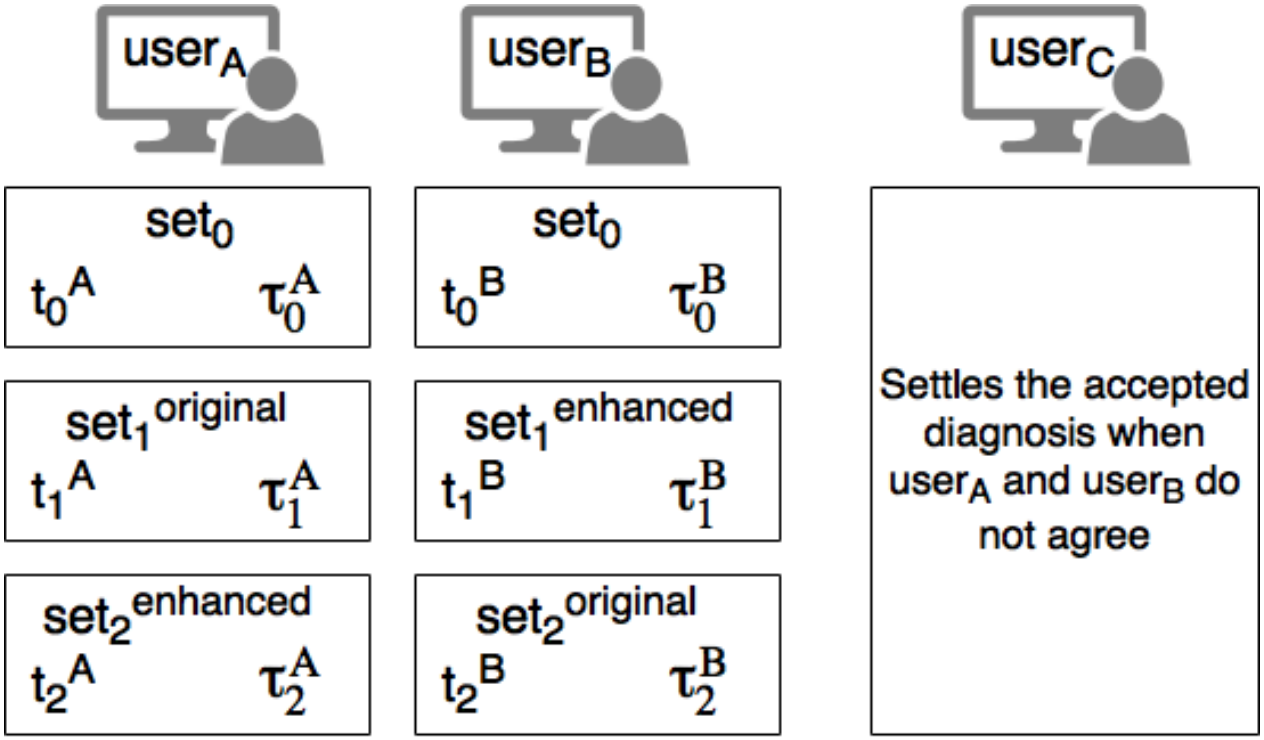


Figure 4.8 Schematic visualization of the experimental protocol.

Our experiment requires three users that we refer to as $user_A$, $user_B$ and $user_C$. We show the previously described sets to $user_A$ and $user_B$ to take our measures. We only ask them to tell for each image if it contains signs of diabetic retinopathy or not, without any more details and without considering any other disease. Since we do not have the ground truth available for the images, we consider that one of the users has made a mistake when both of them do not agree on the diagnosis for a given image. In order to determine which user gave the correct diagnosis, we show the corresponding image to the third user which settles the accepted diagnosis.

More specifically, we show to $user_A$ the images from set_0, $set_1^{original}$ and $set_2^{enhanced}$ and get the average times spent per image t_0^A, t_1^A and t_2^A and the error rates τ_0^A, τ_1^A and τ_2^A on the whole sets. And we show $user_B$ the images from set_0, $set_1^{enhanced}$ and $set_2^{original}$ and get the

average times per image t_0^B, t_1^B and t_2^B and the error rates τ_0^B, τ_1^b and τ_2^B on the whole sets. We show $user_C$ the images where $user_A$ and $user_B$ did not give the same diagnosis to settle which is the accepted answer and thus which user is considered to have made an error.

4.2.2 Results of the user study

The results are reported in Table 4.1 and the details of the measured data can be visualized in Figure 4.9. The first general conclusion that we can draw from these figures is that we can not really demonstrate an impact from our method on the grading speed or accuracy, either positive or negative. Indeed, while the $user_A$ is the fastest on its enhanced set $set_2^{enhanced}$, the $user_B$ is also faster so the speedup can be attributed to the accustomation of the users to our protocol as they where increasingly faster across the sets. And the enhancement does not seem to bear any effect for the $user_B$ as its average time does not differ significantly between the reference set set_0 and its enhanced set $set_1^{enhanced}$. On the error rates as well, no significant change is observed. Overall, these results only let us observe that the users grew accustomed to the task they were asked more than to the images themselves. This observation has been confirmed by their personal feedbacks. According to our analysis of the results, the main factor on the time spent on an image is whether the eye is healthy or not. If the eye presents some signs of a disease, the users spend more time doing their diagnosis and the time spent then far exceeds the estimations of 10 to 20 seconds that we were given for the task. The proportion of images with diabetic retinopathy in each set is approximately the same, meaning that the sets are balanced and the difference of time spent for each of them cannot be explained by a variation of difficulty between them.

Table 4.1 Average time per image and error rates. Measures corresponding to an enhanced set are in bold.

$user_A$		
$t_0^A = 18.1$s	std $= 10.6$s	$\tau_0^A = 0.07$
$t_1^A = 11.6$s	std $= 5.4$s	$\tau_1^A = 0.04$
$\boldsymbol{t_2^A = 8.7}$**s**	**std** $\boldsymbol{= 3.8}$**s**	$\boldsymbol{\tau_2^A = 0.04}$

(a)

$user_B$		
$t_0^B = 22.8$s	std $= 14.6$s	$\tau_0^B = 0.04$
$\boldsymbol{t_1^B = 24.1}$**s**	**std** $\boldsymbol{= 14.4}$**s**	$\boldsymbol{\tau_1^B = 0.05}$
$t_2^B = 15.6$s	std $= 11.5$s	$\tau_2^B = 0.02$

(b)

After the experience, we also asked the users to give us their personal feedback on the experiment and on the images modified with our method. The consensus is that there was a learning curve to the task and that they became more effective at it throughout the experiment. The possibility to zoom on the images to see smaller details was not proposed

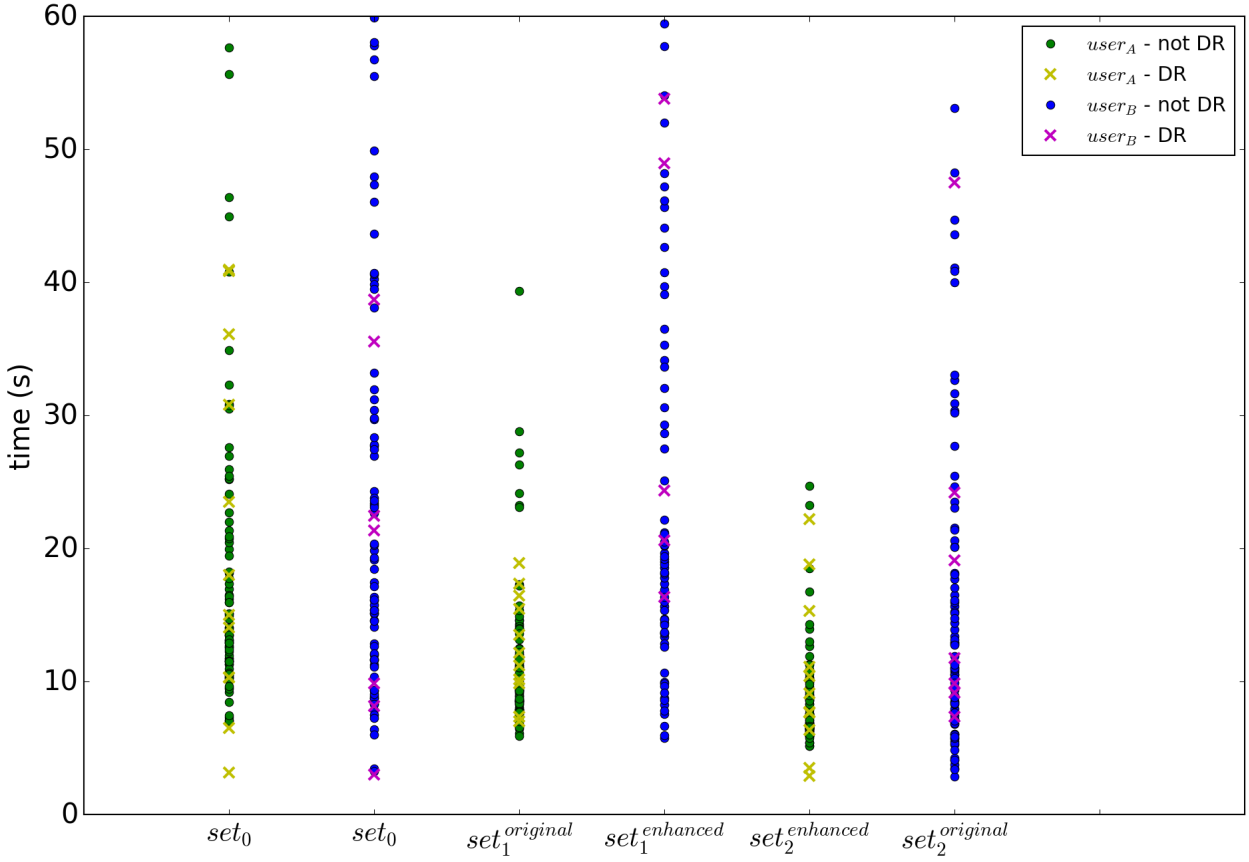


Figure 4.9 User study's measured data. DR stands for diabetic retinopathy.

and $user_A$ noted that it could have helped them in their work as some elements were difficult to see. However he also added that this was less a problem with the enhanced images as the contrast was often better and the lesion more recognizable. But it does not mean that he necessarily preferred the sets with the enhanced images, finding that they sometimes look too different to what they are used to. In particular, he pointed to the contrast of the vessels that is not the same in all the images and the reduced visibility of some details in the optic disc region. These are mainly artifacts that we detail in a following section.

4.2.3 Limitations and biases of the user study

There are some limitations and biases to our user study that were either unavoidable considering our context or that we discovered while putting the experiment into application. This section exposes them and debates their effect on our results.

As the available experts for our experiment were not located close to our offices, the previously

described user study was conducted through a website. Therefore we did not control the conditions in which the users worked on the different sequences and that introduced some variability among the sets as they were not all viewed at the same time and thus not in the same conditions. Some sets of images may have been graded in better conditions than others: with more time at hand, less disturbance from external elements or in better form. We take the perturbation factor into consideration by removing the outliers where the users took anomalously longer than usual, indicating that they have been interrupted in their work and forgot to press the pause button that was proposed for such situations. After analyzing the results and discussing these anomalies with the users, we decided to use 60 seconds as the limit for this threshold, not considering images that exceeded that time as can be seen in Figure 4.9. It is a conservative choice that ensures that we do not take into account any occurrence of such unregistered interruption.

When presenting our results, we commented that the users took longer than what was expected to complete their tasks. This may come from the fact that we asked to only consider the diabetic retinopathy and no other disease in an attempt to simplify and thus speed up their work. In retrospect, this may have been a counter-productive choice and a better option could have been to only ask if the eye was healthy or not, regardless of the specific disease. The $user_A$ shared this opinion in his feedback, noting that it was unusual and thus confusing. Also, since it took them more time than anticipated, it could have made them want to accelerate for the later series. In particular, the $user_B$ was late to complete its last set and may have done it faster for that reason.

The last bias of our user study is that the experiment was designed to be completed by graders whose role is to do a first screening to send to ophthalmologists only the patients that require a further analysis depending on their condition. Due to the lack of availability of such experts, the experience was taken by ophthalmologists whose workflow is different from graders and that may be used to take more time to look at the images. This factor can explain the results for the $user_A$ which has consistently become faster through the experiment as he got used to the task. These elements make it difficult to truly conclude on the data that we collected as too many factors other than our enhancement method may have influenced the results.

4.3 Discussion

In this section, we come back on our objectives in order to assess how our proposed method is successful, critically reviewing our results. We then discuss the limitations of the current work.

As a reminder, we restate our objectives here. Our main objective is to provide a new automatic enhancement method for eye fundus images to help graders visualizing them with better precision while spending less time on each image. Our main hypothesis to attain this goal is that having all the images sharing the same color palette of an image of good quality would help the graders to be faster and more precise. The specific objective is then to provide an automatic color transfer method adapted for the retinal fundus images. It would then be able to distinguish and match the different anatomical elements of the images. We believe that using texture information coupled with a vascular network segmentation to guide the color transfer allows for a correct matching of these elements in terms of colors.

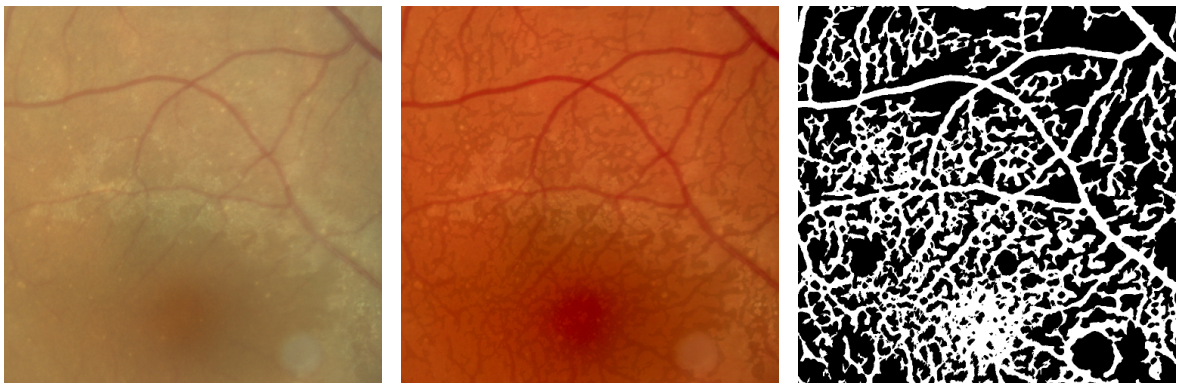
In the previous section dedicated to our results, we presented results that we believe show that our proposed method enhances the images by making them use the same color palette as the carefully selected target. We believe that the variability among images is reduced as all our transferred images share the same colors as can be seen on Figure 4.7 where are presented the color histograms of the images from Figure 4.5. We can see that the color histograms of our results all have a shape close to the histogram of the target image. While the initial result using only texture information in the image gives interesting results, we demonstrated that our proposed changes of using both this texture information and a segmentation of the vascular network improves the correct matching of colors between the vessels. As it has been discussed in the previous section, our experiment did not allow us to observe the expected speed or accuracy increase. However we now have identified the biases that should be taken into consideration for a potential future iteration of this user study.

As we want to improve the screening process, we want our method to be fast enough to not delay significantly the diagnosis. With this constraint in mind, we implement our content based color transfer using C++ and CUDA, leveraging the high parallelization potential of the algorithm. With this implementation, on a workstation with an *Intel i7-6700K*, *16GB RAM* and an *NVIDIA GTX 970* running Ubuntu 16.04 with *CUDA 8.0*, it takes less than a minute to process a 1024x1024 image. We believe that maintaining a computation time of under a minute is a satisfactory constraint as it does not create a significant delay before the image can be observed. A better computer, especially with a higher end GPU, would allow for an increased speed but mostly for bigger images to be processed as the limit is the on-board RAM that on our end could not support higher dimensions images.

4.4 Limitations

Our method has some limitations that we discuss in this section. The first limitation is that we depend on the vascular segmentation and if it contains errors, artifacts may sometimes

occur in the resulting image. This can be seen when a lot of pixels are incorrectly considered as vessels and thus wrongly change the computed colors of the segmented pixels. An example of such a problem can be seen on Figure 4.10. However, in more normal cases we avoid this problem because we use the segmentation along with the texture descriptor and not as a hard replacement of it. Thanks to that, different elements grouped in the vascular segmentation are still considered different.



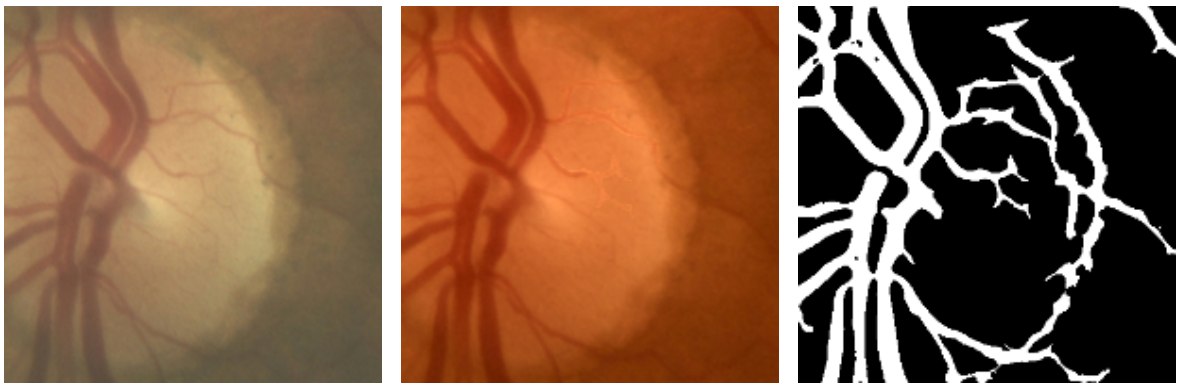
Source | Our result | Vascular network segmentation

Figure 4.10 Color transfer problems due to vascular network segmentation errors. Fundus pixels wrongly segmented as vessels are transferred incorrectly, creating shapes resembling vessels where there are none. (Images courtesy of Dr. Chakor)

Another limitation can be observed in cases where small vessels in the optic disc are in the segmentation along with pixels of the optic disc surrounding it. This causes the vessel to almost disappear in the final image as the texture descriptor characterizes it the same way as the surrounding pixels. The cause of this problem is both the presence of incorrectly segmented pixels and the scale of the texture descriptor that is too large to catch such small structures. An example of this problem can be observed in Figure 4.11. In order to reduce this problem, we could use multiple scales of the region descriptor so that it becomes adapted for such small elements.

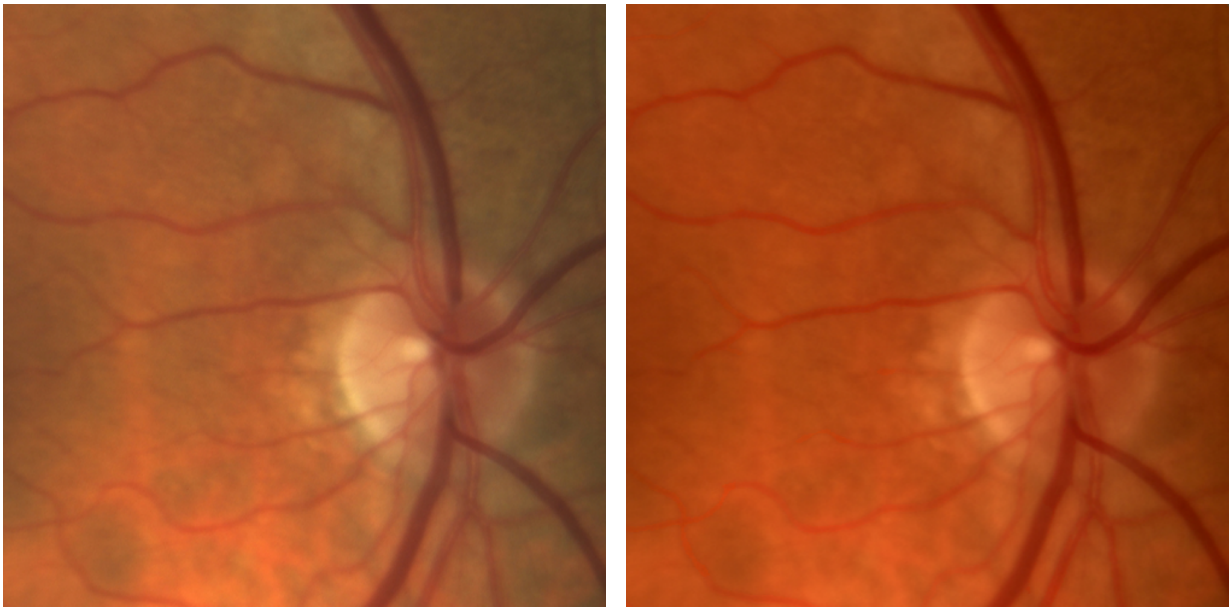
The third limitation is that the proposed algorithm is only designed to enhance the colors of the images. While results show that it sometimes succeeds in reducing problems related to non-uniform illumination and contrast, its focus is not on correcting the artifacts that can be present in the image such as blur or reflection artifacts. Our results on images with these sorts of artifacts are usually not good and this problem is even more noticeable since the vascular segmentation is generally worse in these cases. In Figure 4.12 we observe that blur present on an input image remains on the transferred image.

One last limitation can be observed on the histograms of the produced images and is not



(a) Source (b) Our result (c) Vascular network segmentation

Figure 4.11 Color transfer problems related to small vascular vessels. Small vessels are not correctly represented by the texture descriptor whose scale is too big for them. If fundus pixels surrounding them are in the vascular network segmentation as in (c), they can be transferred with them and thus almost disappear as can be seen on (b). (Images courtesy of Dr. Chakor)

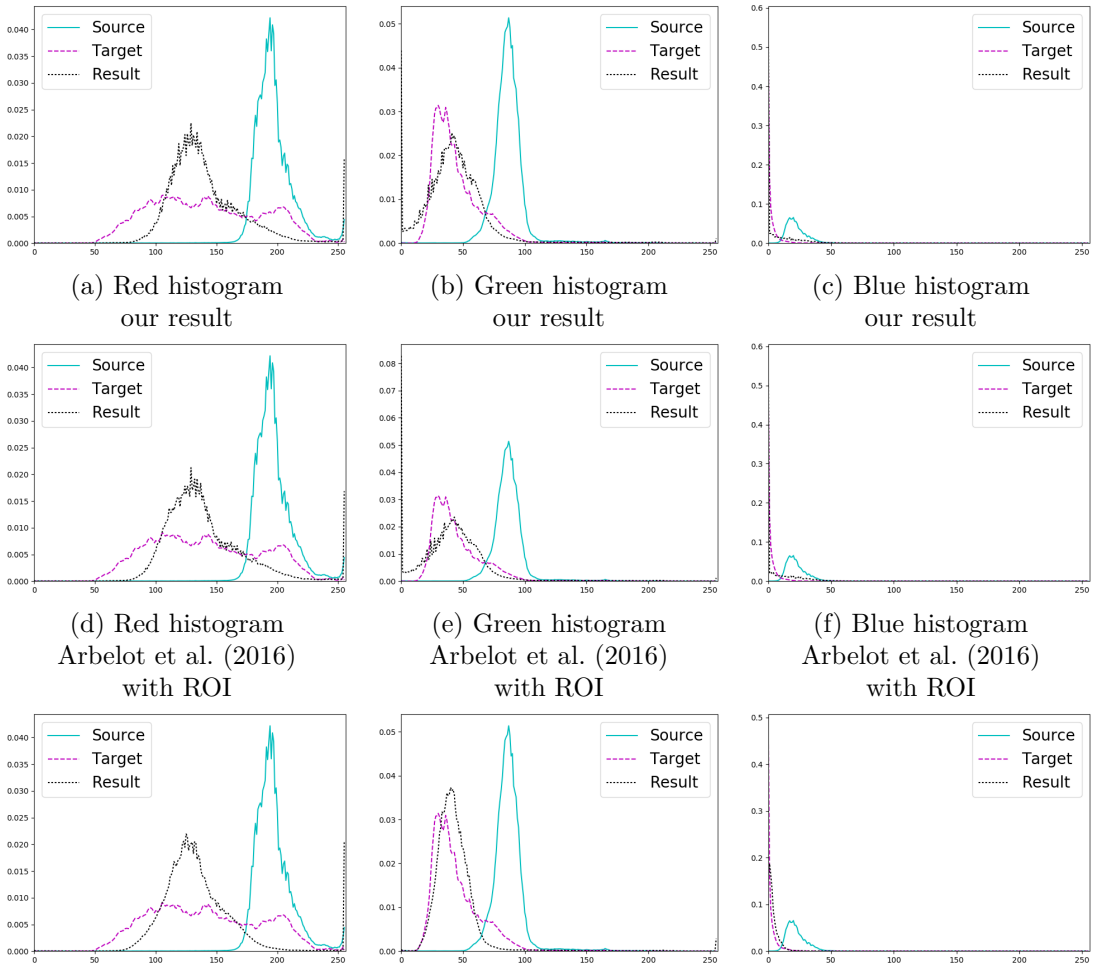


Source Our result

Figure 4.12 Color transfer problems due to blurred input images. Blur, mainly visible in the optic disc area on these images, is not corrected by our method. (Images courtesy of Dr. Chakor)

specific to our proposed method but to the color transfer methods in general. As can be seen in Figure 4.13, there can be a clamping of values to 0 or 255 on the histogram of the resulting image when using a color transfer method. This clamping comes from the

shifting and spreading of the histogram of the source to fit the Gaussian that models the color distribution computed using the colors of the target and our similarity maps. This operation is done on real numbers and may create values that do not belong to $[0, 255]$. This is when the clamping occurs. This typically occurs when, on a given channel, the Gaussian representing the target is closer to one of the limits and wider than the Gaussian representing the source. This phenomenon is observed for all the color transfer we consider, but the pixels whose values are clamped can differ from one method to another as this is dependent on the transformation computed for any given pixel.



(a) Red histogram our result

(b) Green histogram our result

(c) Blue histogram our result

(d) Red histogram Arbelot et al. (2016) with ROI

(e) Green histogram Arbelot et al. (2016) with ROI

(f) Blue histogram Arbelot et al. (2016) with ROI

(g) Red histogram Reinhard et al. (2001) with ROI

(h) Green histogram Reinhard et al. (2001) with ROI

(i) Blue histogram Reinhard et al. (2001) with ROI

Figure 4.13 Values clamping with color transfer methods. (a),(d),(g),(h) Values higher than 255 are clamped to 255. (b),(e) Values lower than 0 are clamped to 0. This happens because the histogram of the output needs to be spread out further than the limiting values to fit the distribution of the target. (Images courtesy of Diagnos Inc.)

CHAPTER 5 CONCLUSION

In this work, we focused on the screening process by human graders of retinal diseases using eye fundus photography images and on how to improve it. These images tend to present a lot of variability especially in terms of colors which can make it harder for the graders to give their diagnosis. We thus proposed a new image enhancement method for the visualization of retinal fundus images that aims at improving the readability of the images as well as reducing the variability among images.

Our method is based on a color transfer between the eye fundus images. We use the texture information in the images along with a segmentation of the vascular networks to guide a local color transfer between a target image with satisfying colors and an input image on which we want to apply these colors. With this information, the anatomical elements of the images are correctly matched and the final colors of their vessels are closer to those of the vessels of the target than with methods proposed by previous works. Our method meets the objective of enhancing the images by using the same good color palette for all images as well as reducing the variability among the images. We conducted an experience that did not allow us to observe an acceleration or an augmentation in precision in the work of the graders because some biases in the experiment prevented us to conclude on the effect of our method. However, we have now identified the elements to take into consideration to conduct a potential future iteration of the user study. Enhancement of readability, especially thanks to better contrast, has also been noted by the graders compared to non modified images.

While our method produces interesting results, it still presents some limitations. Firstly, it does not correct all the existing problems. In particular artifacts such as blur or reflections in the images are not corrected and often induce less convincing results for our method, especially since they can have an impact on the vascular network segmentation. Indeed, we depend on the segmentation of the vessels to improve our results and errors in this step can produce artifacts in the resulting image. This also means that improvements on this subject will have beneficial consequences for our proposed method. The amount of computation of our method is also not negligible and requires powerful hardware to keep it under a minute so that it can be used in the screening process and not delay it too much.

Considering these limitations, further improvements can be considered in the future. While more advanced vascular network segmentation methods could be used to directly improve the results, this may not be the best solution as they usually come with longer computation times. Instead, considering a multi-scale approach with a feature descriptor could be considered to

better recognize the small vessels that are not well described with the texture descriptor. Another development direction could be to replace the transfer on separated decorrelated color channel by a transformation that models the colors as points in a three dimensional space to allow for improved fitting of the models. Future works could also focus on increasing the interactivity with the user. Currently, the only input they can have is the choice of the target image, allowing to use multiple target images or selecting which parts of multiple images they want to use could provide more tools to the users to obtain a satisfying result. The method could also be used to increase the amount of learning data for Machine Learning applications. Indeed, in the same way that using rotations of the images allows to increase the learning data quantity, changing their colors could also serve the same purpose.

BIBLIOGRAPHY

M. D. Abràmoff, M. K. Garvin, et M. Sonka, "Retinal imaging and image analysis", *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.

B. Arbelot, R. Vergne, T. Hurtut, et J. Thollot, "Automatic texture guided color transfer and colorization", dans *Expressive 2016*, 2016.

M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, et F. Durand, "Fast local laplacian filters: Theory and applications", *ACM Transactions on Graphics (TOG)*, vol. 33, no. 5, p. 167, 2014.

S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions", *City*, vol. 1, no. 2, p. 1, 2007.

A. Christodoulidis, T. Hurtut, H. B. Tahar, et F. Cheriet, "A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images", *Computerized Medical Imaging and Graphics*, vol. 52, pp. 28–43, 2016.

D. L. Donoho, "De-noising by soft-thresholding", *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.

H. Faridul, J. Stauder, J. Kervec, et A. Trémeau, "Approximate cross channel color mapping from sparse color correspondences", dans *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 860–867.

S. Ferradans, N. Papadakis, J. Rabin, G. Peyré, et J.-F. Aujol, "Regularized discrete optimal transport", dans *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2013, pp. 428–439.

M. Foracchia, E. Grisan, et A. Ruggeri, "Luminosity and contrast normalization in retinal images", *Medical Image Analysis*, vol. 9, no. 3, pp. 179–190, 2005.

D. Freedman et P. Kisilev, "Object-to-object color transfer: Optimal flows and smsp transformations", dans *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 287–294.

O. Frigo, N. Sabater, V. Demoulin, et P. Hellier, “Optimal transportation for example-guided color transfer”, dans *Asian Conference on Computer Vision*. Springer, 2014, pp. 655–670.

Y. HaCohen, E. Shechtman, D. B. Goldman, et D. Lischinski, “Non-rigid dense correspondence with applications for image enhancement”, *ACM transactions on graphics (TOG)*, vol. 30, no. 4, p. 70, 2011.

Y. Hwang, J.-Y. Lee, I. So Kweon, et S. Joo Kim, “Color transfer using probabilistic moving least squares”, dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3342–3349.

L. Karacan, E. Erdem, et A. Erdem, “Structure-preserving image smoothing via region covariances”, *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 176, 2013.

N. G. Kingsbury, “The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters”, dans *Proc. 8th IEEE DSP workshop*, vol. 8. Utah, 1998, p. 86.

M. M. Nentwich et M. W. Ulbig, “Diabetic retinopathy-ocular complications of diabetes mellitus”, *World journal of diabetes*, vol. 6, no. 3, p. 489, 2015.

U. T. Nguyen, A. Bhuiyan, L. A. Park, et K. Ramamohanarao, “An effective retinal blood vessel segmentation method using multi-scale line detection”, *Pattern recognition*, vol. 46, no. 3, pp. 703–715, 2013.

N. Nida et M. U. G. Khan, “Efficient colorization of medical imaging based on colour transfer method”, *Proceedings of the Pakistan Academy of Sciences: B. Life and Environmental Sciences*, vol. 53, no. 4, pp. 253–261, 2016.

F. Pitié, A. C. Kokaram, et R. Dahyot, “Automated colour grading using colour distribution transfer”, *Computer Vision and Image Understanding*, vol. 107, no. 1, pp. 123–137, 2007.

A. Polesel, G. Ramponi, et V. J. Mathews, “Image enhancement via adaptive unsharp masking”, *IEEE transactions on image processing*, vol. 9, no. 3, pp. 505–510, 2000.

A. Popowicz et B. Smolka, “Bilateral filtering based biomedical image colorization”, *Computational Vision and Medical Image Processing*, pp. 163–169, 2015.

T. Pouli et E. Reinhard, “Progressive color transfer for images of arbitrary dynamic range”, *Computers & Graphics*, vol. 35, no. 1, pp. 67–80, 2011.

E. Reinhard et T. Pouli, “Colour spaces for colour transfer”, dans *International Workshop on Computational Color Imaging*. Springer, 2011, pp. 1–15.

E. Reinhard, M. Adhikhmin, B. Gooch, et P. Shirley, “Color transfer between images”, *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

A. W. Setiawan, T. R. Mengko, O. S. Santoso, et A. B. Suksmono, “Color retinal image enhancement using clahe”, dans *ICT for Smart Society (ICISS), 2013 International Conference on*. IEEE, 2013, pp. 1–3.

J. E. Shaw, R. A. Sicree, et P. Z. Zimmet, “Global estimates of the prevalence of diabetes for 2010 and 2030”, *Diabetes research and clinical practice*, vol. 87, no. 1, pp. 4–14, 2010.

E. Stefánsson, T. Bek, M. Porta, N. Larsen, J. K. Kristinsson, et E. Agardh, “Screening and prevention of diabetic blindness”, *Acta Ophthalmologica Scandinavica*, vol. 78, no. 4, pp. 374–385, 2000.

Y.-W. Tai, J. Jia, et C.-K. Tang, “Local color transfer via probabilistic segmentation by expectation-maximization”, dans *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 747–754.

Z. Wang, F. Wu, et Z. Hu, “Msld: A robust descriptor for line matching”, *Pattern Recognition*, vol. 42, no. 5, pp. 941–953, 2009.

F. Wu, W. Dong, Y. Kong, X. Mei, J.-C. Paul, et X. Zhang, “Content-based colour transfer”, dans *Computer Graphics Forum*, vol. 32, no. 1. Wiley Online Library, 2013, pp. 190–203.

X. Xiao et L. Ma, “Gradient-preserving color transfer”, dans *Computer Graphics Forum*, vol. 28, no. 7. Wiley Online Library, 2009, pp. 1879–1886.

Y. Zhao, L. Wang, W. Jin, et S. Shi, “Colorizing biomedical images based on color transfer”, dans *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on*. IEEE, 2007, pp. 820–823.

APPENDIX A Experimental protocol

A.1 Prerequisites

- 3 experts for the grading of the eye fundus photography images
- 300 eye fundus images

A.2 Preliminary work and notations

- We randomly select 100 images among the 300 to constitute the first set set_0.
- We apply our proposed enhancement method to the remaining 200 images to obtain an ensemble of 400 images (200 original, 200 enhanced).
- We divide this ensemble into two sets of 200 images. Each set is constituted of 100 different eye fundus images, each of them in both available versions (original and enhanced). We name these sets $set1$ and $set2$ and we distinguished within them the subsets corresponding to the original and enhanced images that we note respectively $set_1^{original}$, $set_1^{enhanced}$, $set_2^{original}$ et $set_2^{enhanced}$.
- We refer to the users that are grading the sets as $user_A$ and $user_B$ and to the user that settles which is the accepted answer in case of disagreement between the previous ones as $user_C$.

A.3 Experiment

1. During this experiment, we ask the users to determine for each image whether the eye contains signs of diabetic retinopathy or not, without bringing any further precision and without considering any other illness. In Figure A.2 can be seen the interface of the website that was prepared for the experiment. In particular, it can be noted that a pause button is present to allow the users to temporarily stop the experiment in case of emergency. The pause page that is presented in that case is shown in Figure A.1.
2. We consider that a user has made a mistake when, for a given image, they give a different diagnosis. In this case, $user_C$ settles which is the accepted diagnosis, and then we determine which user is considered to have been mistaken.

3. We present set_0 to both users in order to obtain their reference average time per image (t_0^A and t_0^B) as well as their error rates (τ_0^A and τ_0^B) on regular images.

4. We present $set_1^{original}$ to $user_A$ and $set_1^{enhanced}$ to $user_B$. We measure their average time per image (t_1^A and t_1^B) and error rates (τ_1^A and τ_1^B).

5. We present $set_2^{enhanced}$ to $user_A$ and $set_2^{original}$ to $user_B$. We measure their average time per image (t_2^A and t_2^B) and error rates (τ_2^A and τ_2^B).

Figure A.1 Screenshot of the pause page of the experiment.

Figure A.2 Screenshot of a page of the website for the experiment.

A.4 Results analysis

We want to determine whether our method simplifies the work of the users. We require 2 users to avoid attributing an difference in grading speed or precision to our method when it could come from external factors such as a difference of difficulty among the sets.

In order to be able to attribute an increased speed of grading to our method, we would require to have the following results:

- $t_0^A \approx t_1^A$ and $t_0^B \approx t_2^B$
- $t_2^A < t_0^A$ and $t_1^B < t_0^B$

and results of the following form would mean that our method provoked a decrease in grading speed:

- $t_0^A \geq t_1^A$ and $t_0^B \geq t_2^B$
- $t_2^A > t_0^A$ and $t_1^B > t_0^B$

Any other result would not allow us to conclude that our method has a positive or negative effect on the grading speed as it could come from other external factors.

In order to be able to attribute an improved precision in their diagnosis to our method, we would require the following results:

- $\tau_0^A \approx \tau_1^A$ and $\tau_0^B \approx \tau_2^B$
- $\tau_2^A < \tau_0^A$ and $\tau_1^B < \tau_0^B$

and results of the following form would mean that our method provoked a decrease in grading precision:

- $\tau_0^A \geq \tau_1^A$ and $\tau_0^B \geq \tau_2^B$
- $\tau_2^A > \tau_0^A$ and $\tau_1^B > \tau_0^B$

Any other result would not allow us to conclude that our method has a positive or negative effect on the precision of the users as it could come from other external factors.

A.5 Limitations and biases

After conducting the experiment and gathering the measured data, we identified a few biases in our protocol that prevent us from concluding on the effect of our method on the grading performances. We list them here for reference for any potential future iteration of this protocol:

- The conditions in which the users take the experiment should be controlled to avoid external perturbations to affect the results. Therefore, the experiment should not be implemented into a website that the users can connect to at their own discretion.
- Instead of asking to determine whether the images present signs of diabetic retinopathy, it should simply be asked whether the eye is healthy or not. This will avoid the users to be confused as focusing only on one disease may be more confusing and time consuming.
- This protocol is designed with graders in mind and not ophthalmologists as their respective workflows differ, especially in terms of time spent on a given image. It would then be ideal to conduct the experiment with graders. However if that is not possible, for example because of availability issues, the intended workflow should be carefully explained to the users and a training set should be considered for them to get used to the task.